
mircea bezergheanu

TRAIAN ROTARIU

CURS DE METODE ȘI TEHNICI DE CERCETARE SOCIOLOGICĂ

TRAIAN ROTARIU

C u r s

de

METODE SI TEHNICI DE CERCETARE SOCIOLOGICA

C U P R I N S

-

P R E F A Ţ A

Am adunat în acest volum un set de lecţii din cadrul cursului de Metode şi tehnici de cercetare sociologică, predat studenţilor din anii I şi II ai secţiei de filosofie-istorie, cu speranţa că elaborarea unui text scris va contribui la ameliorarea, pe diferite planuri a activităţii didactice şi va facilita pregătirea studenţilor pentru examen. Necesitatea unui asemenea manual s-a resimţit foarte acut, căci, în ciuda proliferării literaturii sociologice, nu a apărut la noi în ţară nici o lucrare al cărei conţinut să acopere satăsfăcător sfera de probleme tratate în cursul de faţă.

Mărturisesc de la început că ideea centrală care m-a ghidat în structurarea şi selectarea chestiunilor ce vor fi tratate în paginile ce urmează a fost aceea de a oferi viitorilor profesionişti în domeniul ştiinţelor sociale un instrumentar care să-i introducă efectiv în universul cercetării sociologice concrete, să-i ajute în înţelegerea corectă a textelor sociologice de această natură şi chiar să le permită proiectarea şi realizarea unor studii proprii. Din acest motiv, am evitat tratarea unor aspecte filosofice sau istorice ale cercetării sociologice de teren, care se întîlesc frecvent în lucrări similare, ocupînd, nejustificat, un loc mult prea mare în dauna problemelor metodice propriu-zise. Am considerat că asemenea digresiuni nu-şi au locul aici şi pentru faptul că studenţii noştri urmează un curs amplu de "Materialism istoric", unul de "Doctrine sociologice" şi altul de "Epistemologie socială", cursuri care, împreună, oferă o pregătire generală sociologică satisfăcătoare.

Probabil că, sub forma prezentată, textul meu pare încă destul de abstract, presărat cu prea puţine exemplificări concrete. Lucrul se explică prin aceea că am reprodus doar materia predată la curs, ea fiind evident îmbogăţită şi concretizată în seminarii. De altfel, manualul acesta se cere dublat de unul de exerciţii şi probleme, de un ghid practic al cercetării sociologice empirice, pe care sper să-l pot elabora într-un timp cît mai scurt posibil.

Conţinutul cursului este grupat în patru secţiuni principale. Prima cuprinde probleme de natură generală, specificări de termeni şi realizează trecerea în revistă a fazelor generale ale unei cercetări. Am acordat aici o atenţie aparte raportului dintre calitativ şi cantitativ în investigaţia sociologică (cap.3) pentru motivul că, după părerea mea, multe dispute sociologice pe această temă sînt sterile datorită neînţelegerii de către protagonişti a noţiunilor de bază folosite. Este drept că tratarea mea poate lăsa impresia că priveşte doar unele aspecte foarte înguste, foarte specifice ale problemei, în timp ce în literatură atributele de calitativ şi cantitativ sînt folosite în chestiuni de mai largă anvergură, dar rămîn convins că la baza oricărei distincţii între o abordare calitativă şi una cantitativă trebuie să se regăsească în ultimă instanţă, delimitările simple operate aici.

Partea a doua cuprinde trei capitole despre tot atîtea metode: ancheta, experimentul şi analiza documentelor. Textul despre experiment este oarecum diferit de celelalte, în sensul că nu oferă prea multe informaţii practice, ci, evidenţiind esenţa metodei, pune mai degrabă în lumină dificultăţile experimentării în sociologice. Credinţa mea este că sociologia (dacă nu includem în ea şi psihologia socială) rareori are de a face cu situaţii ce pot fi studiate experimental. Am ţinut să subliniez toate exigenţele unui adevărat experiment ştiinţific pentru că nu o dată cercetătorii din ştiinţele sociale îşi etichetează studiile cu calificativul pretenţios de "experimentale", fără o acoperire practică adecvată.

Cu toate că manualele similare apărute în alte ţări poartă pe copertă cel mai adesea titlul de "Metode ale ştiinţelor sociale" (acoperind deci o arie mai vastă decît cea a sociologiei), reflectînd astfel dificultatea unor delimitări şi de metode, nu numai de obiect, între aceste discipline, am evitat o astfel de strategie, în speţă o deschidere prea largă spre psihologia socială, nu atît din motive de concepţie, cît de spaţiu. Aşa se explică lipsa unor lecţii care să trateze despre interviu, testul sociometric, scale de atitudine, eşantioane mici etc., precum şi detaliile practice ale experimentării despre care am amintit mai sus. Dacă tot sînt la capitolul omisiuni să mai amintesc că nici metoda observaţiei nu este tratată într-un capitol special, nu pentru faptul că nu

i-aş acorda importanţa cuvenită, ci deoarece, cred eu, ea nu posedă un sistem structurat de tehnici şi reguli care să justifice un spaţiu de expunere metodică prea larg.

Cum se poate uşor constata, ultimele două părţi, de natură statistică, deţin o pondere însemnată în economia manualului. Am adoptat un astfel de model pentru mai multe raţiuni. In primul rînd sînt convins că fără o abordare de tip statistic a fenomenelor sociale, conţinutul de cunoştinţe dobîndite prin cercetările empirice nu poate fi decît foarte scăzută. Această idee este deseori susţinută în vorbe, dar o analiză atentă a producţiei noastre sociologice lasă să se întrevadă unele lacune serioase în planul punerii ei în practică. Moştenirea unei pregătiri preponderent filosofice se manifestă şi în acest domeniu, unde, cu cîteva excepţii, sociologii noştri se mişcă destul de greoi.

In al doilea rînd, cred că noţiunile simple de statistică introduse în curs sînt necesare nu numai unor viitori sociologi, ci oricărei persoane ce ţine la atingerea unui nivel de cunoştinţe generale cît de cît ridicat. Gîndirea statistică sînt convins că este o componentă esenţială a culturii moderne, iar neglijarea ei de către unii intelectuali de formaţie umanistă nu face decît să-i menţină pe nişte coordonate culturale desuete, depăşite. Mai mult, mi se pare absolut de neconceput ca viitorii profesori de ştiinţe sociale să poată predea satisfăcător elevilor de liceu cunoştinţe social-politice şi economice fără să cunoască noţiuni elementare de statistică. Din acest punct de vedere, dar şi din cel al pregătirii sociologice, mi se pare că volumul de informaţie statistică al manualului este încă nesatisfăcător. S-ar fi impus, dacă spaţiul mi-ar fi permis aceasta, introducerea cel puţin a unor instrumente de analiză a dinamicii fenomenelor sociale.

In al treilea rînd, extensia mai puţin obişnuită a capitolelor statistice într-un curs de "Metode" se explică şi prin convingerea mea că omisiunile la alte teme se pot mult mai simplu recupera, de către cei interesaţi prin lectură individuală. Slabele cunoştinţe matematice ale unei părţi importante a studenţilor filosofi îi privează pe aceştia de posibilitatea consultării aproape a oricărui manual de statistică existent la noi. De aceea m-am străduit să expun aceste probleme într-un limbaj cît mai clar, cît mai

puţin formalizat şi să explic pe larg semnificaţia unor noţiuni şi formule.

Cursul se încheie cu două capitole ce tratează problematica eşantionării. Deşi alegerea eşantionului este o operaţie ce se efectuează în faza pregătitoare a cercetării, introducerea noţiunilor legate de selecţie n-ar fi fost posibilă fără a trece mai întîi prin aspectele statistice din capitolele 7 - 1o.

Lista bibliografică prezentată în final conţine, în principal lucrări metodologice de bază, cărora li se adaugă tratate de statistică generală şi socială, două lucrări generale despre sociologie (Herseni şi Szczepanski) şi alte cîteva lucrări citate în text cu diverse ocazii. Referitor la această listă, încă două precizări. Mai întîi, volumul scrierilor ce ar fi putut fi incluse la bibliografie este foarte mare, dar ţinînd sema de o serie de considerente am ajuns la concluzia că se impune o drastică selecţie a lor. Am ales, aşadar, doar lucrări fundamentale apărute în limba română, de preferinţă cărţi, şi volume străine ce se găsesc în bibliotecile Clujului. Apoi, în textul cursului am evitat pe cît mi-a stat în putinţă introducerea de citate şi aceasta în scopul de a facilita lectura. Desigur, nu am pretenţia că toate ideile ce apar fără trimiteri sînt originale. Dimpotrivă, gradul de originalitate al unui curs, mai ales de profilul celui de faţă, căruia i se cere să sintetizeze o experienţă de cercetare şi cunoaştere bine consolidată, nu poate fi, prin definiţie, decît foarte modest. Aşadar, rostul listei bibliografice de la sfîrşit este dublu: într-un sens , ea indică aria din care au fost sintetizate o parte din ideile lecţiilor (cealaltă parte, mai modestă, provenind din experienţa destul de bogată a autorului în cercetarea empirică) iar, în altul, oferă celor interesaţi posibilitatea lărgirii cunoştinţelor prin consultarea unor surse accesibile. Teribilismul citatelor sau al unei bibliografii nesfîrşite nu-şi avea rostul aici, cum, de altfel, cred că nu şi-l are nici în alte lucrări, care încearcă, în van, prin asemenea metode să-şi ridice statutul de ştiinţificitate, la care aspiră atîta cercetători din domeniul disciplinelor sociale.

PARTEA I

PROBLEME GENERALE

Cap.I. TERMENII PRINCIPALI AI CERCETARII SOCIOLOGICE

In concepţia autorului cursului de faţă, sociologia poate fi şi trebuie inclusă în familia largă a ştiinţelor empirice, în sensul că îndeplineşte - deşi nu în mod perfect şi, cu siguranţă, în măsură mai mică decît alte ştiinţe - cel puţin următoarele patru cerinţe fundamentale ce trebuie avute totodeauna în vedere atunci cînd urmărim să acordăm statutul de ştiinţă unei discipline:

(i) Posedă un obiect empiric propriu de cercetare, decupabil în realitate, fie prin evidenţierea unor entităţi sau fenomene specifice, fie prin precizarea unor modalităţi de abordare, unghiuri de vedere, perspective etc. proprii. Există deci un set de fapte şi fenomene ale universului material a căror descriere, explicare şi previziune, dintr-o anume perspectivă, revin acestei ştiinţe.

(ii) Posedă un sistem conceptual specific, cu ajutorul căruia este "reflectată" realitatea obiectivă şi transformată într-un obiect de cercetare al ştiinţei.

(iii) Discursul ştiinţific este construit prin respectarea legilor generale ale logicii, astfel încît să se poată testa pe această bază consistenţa internă a oricărei deducţii, teorii etc.

(iv) Există modalităţi practice prin care se testează validitatea discursului teoretic, adică tehnici de verificare empirică obiectivă a propoziţiilor ştiinţei.

In cazul special al sociologiei, se cere accentuată importanţa fundamentală pe care acest din urmă aspect o prezintă; peste o atare cerinţă nu se poate trece fără ca disciplina în cauză să se autoexcludă din rîndul ştiinţelor care urmăresc să dea seama de o realitate empirică, trecînd în rîndul altor genuri ale meditaţiei umane ca metafizica, teologia etc. Valabilitatea oricărei propoziţii ştiinţifice trebuie testată, direct sau indirect

prin contactul cu realitatea şi, de asemenea, modalitatea respectivă de testare trebuie să aibă un caracter universal, adică să fie realizabilă de orice om de ştiinţă cu mijloacele şi în condiţiile în care a lucrat autorul aserţiunii emise.

Proiectînd ansamblul exigenţelor mai sus formulate asupra activităţii sociologice, ar fi nevoie, pentru a justifica afirmaţia că ne găsim în faţa unei ştiinţe, să vedem în ce măsură aceste criterii sînt îndeplinite. Altfel spus, s-ar cere delimitat obiectul sociologiei, examinate structura şi funcţionalitatea discursului sociologic şi analizate componenţa şi eficienţa ansamblului de mijloace - domeniul metodologic - prin care sociologia ajunge la descoperirea unor cunoştinţe sau efectuează testarea lor empirică.
Asemenea obiectiv nu poate fi desigur realizat în cadrul restrîns al unui curs ca cel de faţă, care, nici prin profilul său nu urmăreşte aşa ceva, problematica de mai sus constituind obiect de cerßetare în mod deosebit pentru filosofia, logica şi metodologia ştiinţei. In capitolul introductiv prezent voi face doar unele consideraţii cu caracter foarte general şi cîteva precizări terminologice şi de concepţie, indispensabile pentru lecţiile ce vor urma.

1. Obiectul şi specificul sociologiei

A preciza într-o singură frază obiectul unei ştiinţe - cum deseori au pretenţia că o fac autorii de manuale, în deosebi - mi se pare un lucru, în general, dificil (asemenea definiţii avînd aproape întotdeauna forma unui cerc vicios), iar pentru sociologie chiar imposibil, dată fiind aria sa problematică extrem de largă şi caracterul foarte divers şi greu unificabil sub o singură etichetă a aspectelor studiate.

O primă chestiune evidentă este aceea că sociologia este o ştiinţă socială, adică, împreună cu alte discipline ca economia, demografia, psihologia socială, ştiinţele juridice etc. studiază probleme referitoare la viaţa indivizilor în societate, activităţile lor, constituirea şi "funcţionarea" unor instituţii sau colectivităţi ş.a.m.d. Deci spre deosebire de alte ştiinţe ale omului, precum psihologia, disciplinele sociale privesc omul ca o fiinţă socială, interesîndu-se nu atît de trăsăturile individuale



ale acestuia, ci, în special, de raporturile sale cu ceilalţi.

O a doua chestiune ar trebui să fie legată de stabilirea specificului sociologiei în familia mai largă a ştiinţelor sociale. Aici părerile specialiştilor sînt foarte împărţite şi cu greu se pot desprinde cîteva puncte, dacă nu comune tuturor, cel puţin mai larg acceptqte. Mai întîi, sub raportul cuprinderii, aşa cum spunem, sociologia are evident o sferă mult mai amplă, comparativ cu oricare altă disciplină socială, de unde şi evantaiul complex al ramurilor şi subramurilor ei. Incercînd o sistematizare, am putea spune că cercetările sociologice pot avea ca obiect: acţiuni sociale (ca munca, educaţia, propaganda politică etc.), instituţii (şcoala, familia, biserica, partide politice, ş.a.), grupuri sociale,(de la microgrupuri la cele de mari dimensiuni precum clasele sociale), comunităţile umane teritorial determinate(sat, oraş, zonă funcţională etc.) sau o serie de alte fenomene sociale din cele mai diverse (mobilitatea socială, delincvenţa, sinuciderea etc.). Apoi, în ceea ce priveşte specificul de abordare, se poate aprecia că sociologia se caracterizează printr-o perspectivă globală, integratoare,raportînd constant obiectele şi fenomenele cercetate la întregul social, studiindu-le într-un cadru al cauzelor, condiţiilor şi efectelor mai amplu, ceea ce face posibilă şi justifică apariţia unor sociologii de ramură ca cea economică, juridică, a educaţiei ş.a., paralele cu ştiinţele sociale specializate în studierea aceloraşi fenomene, dar dintr-o perspectivă mai îngustă: economia , ştiinţele juridice, pedagogia etc.

Societatea este înţeleasă de marea majoritate a sociologilor ca un sistem complex, alăcuit, la rîndul său , din o serie de subsisteme, care pot integra şi ele altele mai mici ş.a.m.d. Una din sarcinile specifice sociologului constă în analiza cît mai completă a unuia sau altuia dintre aceste subsisteme, adică din punctul de vedere al genezei sale, al modului de funcţionare (structura internă, raporturile dintre elementele lui şi ale subsistemului cu altele de pe acelaşi palier sau de pe cele superioare), al tendinţelor de evoluţie.

Intregul complex de unităţi, fenomene, procese sociale este privit ca avînd o existenţă obiectivă, independentă de subiectul cunoscător. Sociologia ca ştiinţă este posibilă numai în măsura

în care această premisă fundamentală este acceptată, numai dacă viaţa socială prezintă regularităţi, legităţi, numai dacă evenimentele studiate sînt într-un anume fel repetitive, subsumîndu-se unor sisteme cu funcţionare relativ constantă, cu alte cuvinte, numai dacă în universul social există o ordine şi un determinism imanent.

O asemenea idee nu este întotdeauna şi pe deplin acceptată. Anumite orientări filosofice, plecînd de la faptul evident că nici o societate nu există în afara indivizilor umani şi că aceştia din urmă se deosebesc fundamental de elementele materiale ce intră în atenţia ştiinţelor naturii, prin aceea că sînt dotaţi cu conştiinţă, sentimente şi voinţă, care le conferă un anume grad de libertate în acţiunile lor individuale (prin alegerea scopurilor, a mijloacelor de acţiune etc.), în comportamentul lor, în aderenţa sau nonaderenţa la anumite valori etc., susţin că sociologia, ca şi alte discipline socio-umane, s-ar deosebi radical de celelalte ştiinţe şi că modelul naturalist de ştiinţă ar fi inaplicabil aici.

Desigur, nu este locul într-o lucrare ca cea de faţă nici să expun pe larg asemenea concepţii, nici să acord un spaţiu prea mare combaterii lor. Pot şi trebuie doar să-mi justific opţiunea prin cîteva aserţiuni cu caracter de principiu, care mi se par suficiente în acest context. De fapt, argumentele mele nu au nici o pretenţie de originalitate, ele fiind teze de bază ale concepţiei marxiste despre societate, acceptate astăzi de un număr mare de sociologi de cele mai diverse orientări.

Caracterul obiectiv al fenomenelor şi proceselor sociale nu se datorează faptului că acestea s-ar realiza printr-o participare mecanică, automată a oamenilor la ele, că indivizii umani s-ar supune legităţilor sociale aşa cum, de exemplu, pilitura de fier se aşează pe liniile de forţă ale cîmpului magnetic, ci sînt rezultatul acţiunii indivizilor umani concreţi, dotaţi cu voinţă, aspiraţii, trebuinţe, etc. Adică, vreau să spun că anumite generalităţi sînt rezultanteletotumai aleunor astfel de mulţimi de acţiuni individuale, aparent divergente sau haotice,fapt din care rezultă şi o consecinţă metodologică importantă pentru ştiinţele sociale: în cele mai multe cazuri, metodele şi tehnicile de cercetare sînt chemate să producă informaţii capabile să fie supuse

unei analize de tip statistic, pentru a decela regularitățile de acest gen.

Pe de altă parte, chiar dacă acțiunile umane sînt rezultatul unui act de decizie individual, aceasta nu înseamnă actul respectiv dste arbitrar, gratuit. Există scopuri urmărite de majoritatea indivizilor umani, tot așa cum există mijloace mai eficiente pentru atingerea acestor scopuri, ceea ce conferă, în fapt, acțiunilor anumite constante și regularități. Oricît de liber ar părea omul, el este evident împins, cel puțin prin natura sa biologică, să acționeze mai curînd într-o direcție decît în alta.

In sfîrșit , dar nu în ultimul rînd, din momentul în care omul și-a început viața socială, nici o comunitate umană nu poate supraviețui dacă indivizii ce-o compun nu sînt "constrînși" la adoptarea anumitor comportamente, valori, norme etc. comune. Prin urmare, orice societate prezintă regularități de funcționare trans-individuale, independente de voința unuia sau altuia din membrii ce-o compun.

Ansamblul acestor trăsături, care ar putea fi cu ușurință lărgit, fie prin detalierea, fie prin completarea principiilor de mai sus, face posibilă abordarea vieții sociale cu mijloace de cercetare științifică, deci existența unor științe sociale. Această posibilitate se transformă în realitate, deci se ajunge efectiv la știință, dacă activitatea cercetătorului se desfășoară în sensul realizării cerințelor subliniate la începutul capitolului."

2. Stiința sociologică

2.1. Structura științei sociologice

Există astăzi un obicei aproape general de a împărți activitatea științifică în două categorii: una teoretică și alta empirică și de a se face apoi de aici o distincție - deși nu întotdeauna prea convingătoare - între nivelul teoretic și cel empiric al enunțurilor, în cadrul tuturor științelor realului.

Cursul de față nefiind unul de filosofia sau logica științei, nu vom intre în prea multe detalii, mărginindu-ne la cîteva

constatări generale, suficiente pentru a înţelege problematica cercetării sociologice. Fiecare om are, în linii mari, o imagine relativ clară despre ceea ce este o teorie dintr-o ştiinţă oarecare, probabil cea mai familiară şi mai bine conturată, din liceu, fiind teoria mecanicii clasice. De asemenea, latura empirică a cercetării a fost atinsă cu ocazia experimentelor şi observaţiilor în laborator sau pe teren.

In loc de a da o definiţie a teoriei, să-l urmărim pe Lazarsfeld (1971) în descrierea pe care o face acestui nivel al ştiinţei: "Se începe prin formularea unui număr de concepte fundamentale. Unele dintre ele se pretează la măsurări, altele sînt construcţii a priori, a căror validitate rămîne indeterminată la început. Se definesc apoi manipulările la care sînt supuse aceste unităţi elementare şi care vor permite să tragem concluzii noi, pe care le vom verifica la sfîrşit prin observarea faptelor. La un nivel mai elaborat, aceste teorii posedă, în general, două proprietăţi suplimentare: pe de o parte, manipulările şi concluziile sînt exprimate sub o formă matematică şi, pe de alta, demersul tinde să fie 'reductor': proprietăţile gazelor se pot explica prin mişcarea moleculelor, proprietăţile corpurilor prin structura lor atomică" (p.76).

Validitatea unei teorii, cu excepţia celor logico-matematice, nu rezultă numai din rigoarea logică a construcţiei sale, ci cum spuneam, mai ales din confruntarea sa cu realitatea, cu datele experienţei empirice. Gradul de dezvoltare a unei teorii se măsoară adesea prin apropierea structurii sale interne de cea a unui sistem deductiv sau chiar matematic. Puţine sînt însă ştiinţele care au reuşit să-şi construiască un asemenea sistem perfect coerent. Din ignoranţă, este destul de răspîndită ideea după care ştiinţele naturii în totalitate (sau în marea lor majoritate) ar poseda o astfel de bază teoretică, pe cînd cele socio-umane ar fi lipsite de ea. De aici rezultă firesc ideea că acestea din urmă sînt subdezvoltate, iar pînă la a spune că ele nici n-ar putea ajunge la rigoarea primelor nu este decît un pas. După cunoştinţele mele, cazurile fericite de ştiinţe ale naturii cu o structură axiomatico- deductivă sînt însă mai degrabă excepţii decît regula .

Adevărul este, cred, că foarte multe ştiinţe, atît din domeniul fenomenelor naturii cît şi din cel al vieţii sociale, folosesc mai ales scheme explicative, care nu derivă dintr-o teorie bine pusă la punct. Cel mai adesea, aceste scheme nu pot fi nici perfect coordonate şi coroborate, astfel încît, pe baza lor, să se poată ajunge la construcţii teoretice de mare generalitate. Desigur, gradul de rigurozitate al discursului nu este acelaşi , chiar şi în rîndul celor mai puţin închegate. Un astfel de grad ar putea fi evaluat tocmai prin umrărirea măsurii în care schemele teoretice explicative se leagă unele de celelalte: la un pol am avea discursurile axiomatizate, la celălalt cele care produc scheme explicative ad-hoc, pentru fiecare fenomen studiat. Intre cele două extreme este un spaţiu de joc enorm de mare, unde fiecare ştiinţă îşi găseşte un loc pe o poziţie mai înaltă sau mai joasă. Atunci cînd un număr de scheme explicative se "coagulează" în jurul unui nucleu de propoziţii de bază , poate lua naştere un subsistem teoretic explicativ al unui fragment de realitate. Se nasc astfel, ceea ce Merton a denumit cu o expresie devenită deja clasică în sociologie, teoriile de rang mediu .

Dacă, referitor la structura ştiinţei, distingerea nivelului teoretic de cel empiric nu este o operaţie prea uşoară (ba chiar există, cum spuneam, temeiuri de a crede că nici nu este posibilă o asemenea distincţie), în activitatea practică a cercetătorului este mai vizibilă diferenţa dintre acţiunea de construcţie teoretică şi cea de investigare empirică a realităţii.

Urmărind raţionamentul lui Lazarsfeld (1971), activitatea teoretică în sociologie poate fi sistematizată prin operaţiile următoare:

"-stabilirea de scheme clasificatorii precise,

- formularea de concepte complexe care orientează observatorul spre fapte interesante,

- formularea de probleme de cercetare avînd o mare importanţă din punctul de vedere al societăţii,

- formularea de idei generale asupra manierei în care schimbările actuale se produc sau pot fi provocate,

- previziuni fondate pe descoperiri empirice încă neverificate (ipoteze).

- punerea în relaţie a faptelor empirice cu altele, ipotetice sau deja verificate (interpretarea)" (pp.76-77).

Activitatea la nivel empiric constă, esenţialmente, în abordarea concretă a realităţii sociale cu ajutorul metodelor şi tehnicilor de investigaţie empirică. Vom considera deci ca fiind de natură empirică operaţii ca experimentul, anchetele pe bază de chestionar, analiza unor documente a căror informaţie nu este sociologic sistematizată, prelucrarea unor date statistice primare etc.

Raportul de forţe dintre cele două genuri de activităţi fundamentale în sociologie, cea teoretică şi cea empirică, a fost adesea schimbător. Dacă în secolul trecut şi pînă la primul război mondial , se poate afirma că a predominat abordarea teoretică (a fost epoca marilor sisteme sociologice, de regulă speculative), în perioada interbelică, cu deosebire în Statele Unite, a luat un avînd cu totul remarcabil cercetarea empirică. Ar fi aci de subliniat că toate tehnicile de cercetare concretă au fost preluate de sociologii americani din arsenalul europenilor, care le-au utilizat doar sporadic în perioada anterioară şi care, cu puţin excepţii (printre ele numărîndu-se şi şcoala sociologică de la Bucureşti a profesorului D.Gusti) au manifestat aceeaşi atitudine prudentă în cercetarea empirică pînă la al doilea război. Spre sfîrşitul perioadei interbelice, sociologia empirică americană depăşise faza de amatorism, astfel că această formă de cercetare a cunoscut rapid o largă răspîndire în toate ţările occidentale. Reacţiile împotriva empirismului excesiv n-au întîrziat să apară, mai cu seamă pe bătrînul continent, şi aşa se face că din anii '6o sociologia caută să găsească justul echilibru între cele două tipuri de activităţi.

Excesele, atît speculative cît şi cele empiriste, au dus astăzi majoritatea sociologiclor în situaţia de a înţelege mai clar:

2.2. Raportul dintre teoretic şi empiric în sociologie

O primă achiziţie, aş spune definitivă, a sociologiei contemporane constă în dobîndirea conştiinţei sale de a aspira la statutul de ştiinţă pozitivă şi, prin urmare, în recunoaşterea necesităţii de a se dezvolta concomitent prin consolidarea edificiu-

lui său teoretic, prin îmbogăţirea şi verificarea , forjarea în practică, a arsenalului metodologic şi prin acumularea neîncetată de informaţii asupra realităţii sociale.

O a doua achiziţie a sociologiei ultimelor decenii ar putea fi numită "împăcarea" cercetării empirice cu cea teoretică. Desigur că nici în perioada sociologiei aşa-zis speculative şi nici în cea preponderent empirică nu a existat un veritabil "divorţ" între cele două sfere de activitate, dar este clar că, exceptînd anumite cazuri particulare, ce nu fac decît să întărească regula, accentele erau puse prea puternic pe una sau alta din laturi. Actualmente se recunoaşte şi se subliniază atît importanţa teoriei pentru cercetarea concretă, cît şi influenţa de necontestat pe care investigaţia empirică o are în dezvoltarea teoriei.

In legătură cu primul aspect, aş menţiona, în special, ideile sociologului francez Pierre Bourdieu, unul dintre campionii luptei împotriva empirismului plat. In lucrarea Le métier de sociologue (1968), scrisă împreună cu doi colaboratori, din care s-a reprodus un fragment şi în Antologia sociologiei franceze contemporane (1971), după care citez, este spulberată ideea unei cercetări empirice pure, care să nu facă apel la nici un element teoretic, deci neutre faţă de o teorie sau o concepţie despre viaţa socială. In acest sens se precizează că "orice practică ştiinţifică, chiar atunci şi mai ales cînd se revendică orbeşte de la empirismul cel mai orb, angajează presupuneri teoretice şi că sociologul nu are de ales decît între interogări inconştiente, deci necontrolate şi incoerente, şi un corp de ipoteze construite metodic în vederea verificării experimentale" (p.192).

Dacă, aşa cum spun sociologii francezi , "măsura şi instrumentele de măsură şi, mai general, toate operaţiile practicii sociologice, de la elaborarea chestionarelor şi a codificării pînă la analiza statistică, sînt tot atîtea teorii în act luate ca proceduri de construcţie, conştiente sau inconştiente, de fapte şi relaţii între fapte" (p.192), dacă, zic, lucrurile stau în acest fel, atunci devine evident rolul teoriei în toate etapele cercetării empirice, de la alegerea problemei de studiat şi pînă la redactarea raportului final de cercetare. Definirea operaţională a noţiunilor, construcţia dimensiunilor şi a indicatorilor empirici

pentru acestea, numărul şi natura întrebărilor dintr-un chestionar, modalităţile de prelucrare a datelor şi chiar schema de eşantionare folosită implică o anumită concepţie despre realitatea studiată.

Mai cu seamă rezultatele finale ale unei cercetări nu apar niciodată sub formă empirică pură. Toate datele statistice, de pildă, se cer interpretate, şi asta înseamnă încă o dată - şi în modul cel mai esenţial - apelul la o teorie. Celebra butadă după care cu ajutorul statisticii se poate demonstra orice vine să ilustreze o stare de fapt; atunci cînd teoria lipseşte din cercetare, aceleaşi date pot fi uneori prelucrate şi prezentate ca să susţină concluzii contradictorii, nu pentru că datele ar fi inconsistente, ci pentru că modalităţile de a le prelucra sînt în funcţie de concepţii diferite despre fenomenele respective.

Concluzia firească ce se impune este că dacă orice cercetare empirică implică astfel de presupoziţii teoretice, este preferabil ca ele să fie conştientizate şi explicite înaintea travaliului practic. Chiar dacă sociologia nu beneficiază de teorii suficient de elaborare, este posibilă constituirea, în faza de pregătire a cercetărilor empirice, cel puţin a unui corp de ipoteze, care să ghideze demersul practic şi care să ajute la precizarea conceptelor de lucru. Desigur,că, pe de altă parte, o teorie relativ închegată are un rol fundamental pentru cercetările concrete. In cadrul ştiinţelor ce posedă practic astfel de teorii, cercetarea empirică este în proporţie covîrşitoare nu numai subordonată prin concepte acestora, dar şi <u>impusă</u> în actele sale concrete, dirijată, canalizată spre problemele ridicate de teorii. In cazul sociologului însă, studiile empirice se efectuează rareori în intenţia de testare a unei teorii iar alegerea problemelor se datorează precumpănitor unor factori extrateoretici.

Referitor la rolul cercetării empirice în dezvoltarea teoriilor voi folosi un text tradus în româneşte al sociologului american R.K.Merton (1972), ale cărui idei acoperă satisfăcător această problemă şi care se pot rezuma în felul următor.

După părerea lui Merton, a existat o oarecare unilateralitate în abordarea chestiunii de faţă, în sensul că discuţiile sociologice au atribuit drept unică funcţie majoră cercetării empirice

testarea sau verificarea ipotezelor (p.285). Desigur, nu este niciodată de prisos să se insiste asupra acestei funcţii a studiului concret, ea fiind esenţială în orice ştiinţă. Nici o teorie asupra realităţii nu se poate justifica fără contactul cu realitatea; verificarea teoriei înseamnă esenţialmente testarea empirică a propoziţiilor ei. Dar, cum precizează Merton , "cercetarea empirică nu se limitează la rolul pasiv de a verifica şi testa teoria" (p.285), şi aceasta, îmi permit să adaug, cu atît mai mult în domeniul unei ştiinţe în curs de constituire, precum sociologia, unde schemele teoretice suferă, cum am mai spus, de o mare fragilitate. După autorul american, patru ar fi funcţiile majore ale cercetării empirice în dezvoltarea teoriei:

a) Ea iniţiază teoria (fenomenul de serendipitate). Serendipitatea este un termen prin care Merton desemnează descoperirea unui fapt neaşteptat, aberant şi capital. Neaşteptat, pentru că el apare într-o cercetare în care nu se urmăreşte aşa ceva; aberant, deoarece faptul respectiv este inconsistent cu teoria; capital, fiindcă descoperirea are consecinţe importante asupra teoriei.

b) Ea reformulează (reconstruieşte) teoria. Cercetarea unor date noi, neglijate pînă atunci, care nu pot fi explicate cu schemele teoretice existente, conduce la elaborarea unor noi scheme explicative.

c) Ea orientează interesul teoretic . Metodele noi de cercetare empirică pot duce la noi direcţii de interes teoretic. Este deci înţeleasă aici mai degrabă o influenţă a instrumentarului de cercetare asupra teoriei. Cred că este surprins în mod judicios faptul că există o interdependenţă între domeniul teoretic şi cel metodologic; nu numai faptele determină teoria, dar şi inventarea unor tehnici sau rafinarea altora cunoscute pot orienta interesul teoretic asupra unor domenii neglijate sau, eventual, abandonate speculaţiei pure.

d) Ea ajută la clarificarea conceptelor. Dacă într-un discurs speculativ cu pretenţii teoretice se pot folosi termeni mai mult sau mai puţin definiţi, prin implicarea conceptelor în cercetarea empirică acestea trebuie să dobîndească claritatea necesară identificării obiectelor reale ale căror proprietăţi se sub-

sumează conţinutului noţiunilor respective.

2.3. Statutul ştiinţific al sociologiei

Aşa cu, am mai remarcat, nivelul teoretic al sociologiei nu străluceşte prin virtuţi demne de invidiat; de aici şi o modestă poziţie ştiinţifică a sociologiei. O privire superficială, aruncată fie din interiorul şi mai cu seamă din exteriorul acestei discipline va surprinde o stare ce, în termeni foarte succinţi şi poate prea brutali, ar putea fi descrisă în felul următor. La nivelul teoriei generale, se constată o luptă ascuţită între diferitele orientări, în ciuda faptului că actualmente în occident se manifestă o clară penurie de asemenea teorii care ar putea concura marxismul. In sociologia marxistă, dezvoltările teoretice sînt relativ timide, ele reducîndu-se în bună măsură la asimilarea unor concepte şi scheme explicative utilizate cu mai mult sau mai puţin succes în alte cînpuri teoretice. La nivelul teoriilor de rang mediu sau al schemelor explicative observăm o diversitate deconcertantă iar la cel al cercetării empirice s-ar putea vorbi chiar, dacă termenul nu e prea tare, de o adevărată junglă de date, de scheme explicative ad-hoc şi de instrumente de investigaţie.

O analiză mai profundă poate pune totuşi în evidenţă anumite principii şi părţi ce fac posibilă o unificare între diferite teorii, iar, la nivelul investigaţiei empirice, rezultate similare, parţial verificabile. Vorbind despre aceste aspecte la Congresul de sociologie de la Evian, în 1966, reputatul sociolog american P.Sorokin (1967) asemuia sociologia cu un Ianus cu două feţe, din care prima (lipsa de unitate) este mai uşor de remarcat, din următoarele motive:

- se pleacă de la premisa că în alte ştiinţe nu ar fi diversitate şi chiar discordanţă a teoriilor;
- în sociologie apare cu pregnanţă problema limbajului foarte diferit, care însă la o cercetare mai atentă se dezvăluie a fi mult mai uniform decît în aparenţă (mulţi termeni se suprapun, total sau parţial);
- sociologii desconsideră adesea "principiul limitei", adică au tendinţa de a-şi extinde propoziţiile dincolo de domeniul lor de validitate;

- sociologii neglijează aspectul multidimensional al realităţii sociale şi exagerează o dimensiune.

In ciuda faptului că în ultimele decenii s-au făcut progrese însemnate spre o mai mare rigurozitate, atît în construcţiile teoretice, cît şi la nivelul cercetării empirice, gradul de consistenţă al sociologiei se cere încă substanţial consolidat. Rămîn, de pildă, valabile cuvintele spuse de un alt mare sociolog, E.Shils (1967), la acelaşi congres de la Evian; după a sa părere, una din deficienţele majore ale sociologiei rezidă în faptul că teoria nu a ajuns încă suficient de concretă pentru a veni în contact cu datele reale şi, de aceea, nu e în măsură să-şi încorporeze, la rîndul ei, rezultatele analizei acestor date.

Care ar fi, în fond, explicaţia faptului că sociologia se găseşte într-o asemenea stare? Incercările explicative n-au lipsit, ba dimpotrivă sînt atît de numeroase încît numai o analiză a lor ar umple probabil un tom. Fără pretenţia de a fi sistematizat corect şi, cu atît mai puţin, de a fi inventariat complet toate ideile emise în legătură cu această problemă, voi enumera unele din cauzele citate în literatura filosofico-metodologică pentru a explica statutul ştiinţific mai modest al sociologiei:

a) complexitatea fenomenelor sociale, ceea ce înseamnă că orice analiză de efectuat şi orice legitate de stabilit trebuie să implice un număr însemnat de variabile;

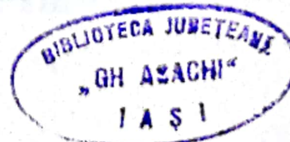
b) legat de aceasta, dificultatea izolării unor aspecte autonome;

c) efemeritatea multor fapte sociale, schimbările fiind evident mai rapide în societate decît în majoritatea celorlalte zone ale existenţei;

d) lipsa unui "contur" precis al anumitor fapte sociale, mai ales cele ce ţin de conştiinţa socială;

e) existenţa indivizilor umani în spatele oricăror fapte sau acţiuni sociale, ceea ce impune includerea în explicaţii, dacă nu a întregului univers psihic al acestora, cel puţin a unor elemente de bază, precum <u>finalitatea</u> acţiunilor individuale;

f) observatorul se găseşte "scufundat" în realitatea socială, fiind copleşit de aceasta (paradoxul ochiului care se autocontemplă);



g) dificultăţi legate de măsurarea multora (dacă nu a majorităţii) dintre aspectele vieţii sociale;

h) dificultăţi legate de posibilităţile efective de observare şi, mai ales, de experimentare;

i) lipsa unor instrumente neutre, ca în fizică, de exemplu, ceea ce face ca în sociologie observatorul să fie, într-un fel, el însuşi şi instrument, de unde rezultă atît imperfecţiunea observaţiei, cît şi un adaos de subiectivitate;

j) condiţionarea socială şi culturală a cercetătorului;

k) legătura imediată a rezultatelor cercetării sociologice cu viaţa de zi cu zi a oamenilor, de unde posibile presiuni extra-ştiinţifice şi manipulări din partea factorilor ce dispun de o asemenea putere;

l) lipsa unor informaţii primare consistente, obiective şi acoperind intervale temporale lungi;

m) existenţa unei "sociologii vulgare";

n) informaţia fiind luată adesea de la indivizi umani nu prin observaţie externă, neutră, ci prin interogare, intervievare, sociologul se ciocneşte de un ecran interpus între el şi realitate, fie în sensul că poate lua drept adevărate informaţii false (este minţit, pur şi simplu), fie că trebuie să realizeze o interpretare mediată a realităţii. Este aici problema surprinsă foarte plastic de Borudieu şi colaboratorii (1971), cînd afirmă că "este un blestem ca ştiinţele umane să aibă de a face cu un obiect care vorbeşte..." (p.192).

Cum se vede, multe dintre motivele invocate sînt legate unele de altele şi s-ar cere, desigur, o mai amplă explicitare a lor şi o analiză ceva mai aprofundată, dar spaţiul nu-mi permite acest lucru.

Ceea ce doresc să mai adaug este faptul că, pe lîngă factorii amintiţi, factori intrinseci care ţin de natura însăşi a sociologiei, există şi alţii, exteriori, legaţi de funcţia şi rolul unei ştiinţe în societate, avînd o influenţă de netăgăduit asupra acesteia. Se ştie că (şi lucrul este mai evident în societăţile moderne) activitatea ştiinţifică a căpătat pe măsura trecerii timpului un caracter de interes social tot mai important. Astăzi este din ce în ce mai greu de a întîlni cercetători ai



ştiinţelor empirice care să lucreze singuri, izolaţi şi pe cont propriu. Munca în echipe, un anumit nivel minim de dotare tehnică, chiar şi pentru ştiinţele sociale, pun în evidenţă dependenţa tot mai pronunţată a cercetării de factorii de decizie, de disponibilităţile bugetare pe care o societate sau alta le are pentru dezvoltarea ştiinţei. Dacă lucrurile sau aşa, este limpede că pentru gospodărirea acestor mijloace - forţamente limitate - se impune stabilirea anumitor priorităţi îm alocarea lor. Or, trebuie s-o recunoaştem, ştiinţele sociale sînt net handicapate în raport cu celelalte în accesul la asemenea mijloace, cel puţin din trei motive fundamentale:

a) Intr-o anumită fază de dezvoltare a societăţii, prioritară apare dobîndirea unor cunoştinţe şi mijloace care să permită o rapidă dezvoltare economică. La început deci, cercetarea ştiinţifică este orientată spre acele domenii care permit creşterea productivităţii muncii, dezvoltarea bazei de materii prime etc. prin mijloace pur tehnice. Abia realizată o anumită fază în acest proces, se vor face simţite ameliorările pe care le pot aduce cercetările socio-umane. Aşa de pildă, mi se pare cu totul evident că nici o organizare ştiinţifică a producţiei într-un atelier nu poate ridica productivitatea muncii în aceeaşi proporţie ca introducerea unei maşini ce efectuează într-un minut munca unui om pe timp de o zi, să zicem.

b) Dat fiind statutul lor ştiinţific şi natura fenomenelor abordate, ştiinţele sociale nu pot oferi soluţii nici atît de prompte şi nici atît de sigure ca cele tehnice. Introducerea unei maşini noi într-o întreprindere - ca să reiau acelaşi exemplu - poate fi uşor acompaniată de o serie de calcule de eficienţă, tocmai pentru că asemenea calcule privesc un domeniu foarte strîmt de aspecte (economice, de cele mai multe ori). Disciplinele sociale - sociologia mai ales - se preocupă de fenomene foarte complexe şi,ca atare, gradul de calculabilitate al consecinţelor unei intervenţii sociale este extrem de redus.

c) Există o serie de reacţii negative faţă de ştiinţele sociale, mult mai evidente şi chiar mai virulente decît faţă de altele şi mult mai diversificate din perspectiva statutului social al

celor care le manifestă. Este clar, mai întîi, că cercetarea socială nu este totdeauna comodă pentru anumite persoane cu funcţii de conducere. Un sociolog care "scormoneşte" în relaţiile interpersonale sau în alte probleme ale unei intreprinderi trebuie să beneficieze de concursul unor cadre de conducere extrem de bine intenţionate şi cu o conştiinţă deosebită. Teama multora dintre astfel de oameni de a nu-şi pierde funcţia, dacă cercetarea sociologică dă la iveală anumite aspecte negative, este un lucru real de care trebuie să se ţină întotdeauna seama.

Dar nu numai din atari motive pragmatice pot apare reacţii negative la adresa cercetătorilor sociologice. Există o credinţă destul de răspîndită în rîndul oamenilor că explicarea fenomenelor sociale, prin faptul că sînt mai uşor perceptibile, este o treabă la îndemîna oricui. Parafrazînd o cunoscută expresie, aş putea spune că toţi indivizii cred că sînt născuţi sociologi; fiecare are o explicaţie, mai mult sau mai puţin personală, pentru orice fapt, de la întîrzierea unui autobuz aşteptat în staţie şi pînă la izbucnirea unui conflict armat între două state; fiecare face, este dispus - şi într-un fel constrîns de însăşi viaţa lui în societate - să facă previziuni asupra consecinţelor actelor sale sau ale altora, fiecare este gata să ofere soluţii pentru îndreptarea tuturor neajunsurilor.

Si nu este numai atît. O bună categorie de intelectuali, mai cu seamă din rîndurile literaţilor, consideră că rezultatele la care ajung ştiinţele sociale sînt minore în comparaţie cu complexitatea fenomenelor studiate şi că literatura sau arta sînt capabile să dezvăluie lucruri mult mai profunde despre om şi societate. Nu este locul să analizeze aici raportul dintre cele două domenii de activitate intelectuală - ştiinţele sociale şi arta (mai ales literatura) - nici nu doresc să neg rolul pe care îl poate avea în dezvăluirea unor imagini ale realităţii sociale şi nici nu ţin să analizez cauzele disputelor numeroase dintre reprezentanţii celor două sfere de activitate. Aş afirma doar că reproşurile făcute disciplinelor sociale sînt în mare măsură nedrepte, după cum inoportună este contrapunerea celor două genuri de activitate, care nu sînt deloc concurente, ci complementare şi nici o cultură modernă nu-şi mai poate permite să se dispenseze de una din ele.

In fine, mai există o reacţie negativă vizavi de ştiinţele sociale, pe temeiul faptului că acestea, prin rezultatele consistente la care ajung - sau ar ajunge - pot servi pentru manipulare ori condiţionarea oamenilor. Reproşul este copilăresc; este ca şi cum ai vrea să interzici fizica deoarece dezvoltarea ei a condus la fabricarea armelor nucleare. Stiinţa în sine nu este nici bună nici rea, numai utilizarea ei poate fi favorabilă sau defavorabilă omenirii.

3. Domeniul metodologic

Intreg cursul de faţă va încerca în continuare să descrie unele aspecte ale acestui domeniu. In prealabil, se cuvin făcute două lucruri: definirea unor noţiuni sau, mai degrabă, precizarea sensului în care ele vor fi folosite aici şi enunţarea cîtorva idei despre valoarea reflexiei metodologice.

3.1. Termenii

Diversitatea terminologică în domeniul de care ne ocupăm e de-a dreptul deconcertantă. Termeni ca metode, tehnici, proceduri, procedee, instrumente etc. se folosesc în accepţiuni foarte diferite, încît este greu să mai găseşti ceva comun în acestea. Mai ales termenul de metodă este utilizat pînă la deplina uzare şi pierderea oricărui conţinut precis. Se vorbeşte de metoda dialectică, de metoda experimentală, de metode matematice sau statistice, de metode de eşantionare ş.a.m.d. Este greu să se pună ordine în acest haos printr-o convenţie unanim acceptată şi este dificil şi pentru acelaşi om să fie absolut consecvent în folosirea termenilor respectivi. Nu-mi propun deci o clarificare a limbajului, ci voi face doar cîteva precizări terminologice pe care voi încerca să le respect, pe cît posibil, pînă la sfîrşitul cursului.

Prin domeniu metodologic înţeleg totalitatea mijloacelor, modalităţilor prin care o ştiinţă parvine la rezultatele sale. Aceste mijloace pot fi folosite fie în domeniul cercetării teoretice, fie în cel empiric şi este clar că cele două moduri de activitate ştiinţifică uzează de mijloace diferite . In cursul de faţă,

voi prezenta unele din problemele tehnice ale cercetării empirice, care ajută la obținerea informației sociologice și la o prelucrare sumară a acestora.

Termenul de metodologie are o gamă largă de utilizări. In ceea ce mă privește, îi voi atribui un singur sens, și anume cel care derivă din etimologia cuvîntului: teorie (discurs) asupra metodei. Mai exact, metodologia este o ramură științifico-filosofică ce are ca obiect de studiu domeniul metodologic al științei, urmărind stabilirea unor legități referitoare la valoarea de cunoaștere a diferitelor metode de cercetare. Vom avea deci o metodologie generală, corespunzînd domeniului metodologic al tuturor științelor și metodologii particulare, pentru fiecare știință în parte sau pentru grupuri de științe apropiate. Vom putea vorbi astfel de o metodologie sociologică sau a științelor socio-umane.

Sensul acesta este destul de apropiat de cel propus de Lazarsfeld (1970), conform căruia metodologia nu este înțeleasă ca o disciplină normativă - cum se întîmplă de multe ori - chemată să prezinte reguli de folosire a diferitelor mijloace de cercetare, ci una pozitivă, avînd drept sarcină de bază urmărirea sociologului în activitatea sa efectivă. Destul de apropiat, dar nu identic, deoarece o serie de funcțiuni pe care Lazarsfeld le atribuie metodologiei aparțin de fapt de domeniul teoretic al fiecărei științe. Dar iată care ar fi, după părerea sociologului american, temele majore ale preocupărilor metodologilor:

a) delimitarea obiectelor, în sensul clarificării aspectelor decupate de sociologi din problematica vastă studiată;

b) clarificarea sensului termenilor folosiți;

c) explicarea tehnicilor de cercetare;

d) punerea în relație a diferitelor tehnici de cercetare;

e) sistematizarea rezultatelor empirice;

f) formalizarea raționamentului (pp.258-266)

Al treilea termen fundamental ce se cuvine precizat este cel de metodă. Voi reține două dintre accepțiunile sale cele mai răspîndite. Mai întîi, prin metodă se va înțelege o modalitate ge-

nerală de abordare a realității, un principiu fundamental ce ghidează întreaga desfășurare a unei cercetări. Astfel, se poate vorbi de o metodă structuralistă, funcționalistă, dialectivă, istorică, genetică, monografică, interdisciplinară etc. Se vede din înșiruirea de mai sus cît de eterogene sînt categoriile folosite, în sensul că ele nu sînt nici exclusive, nici de aceeași natură și nici de aceeași anvergură; unele din aceste tipuri de abordări derivă din concepția filosofică, altele din teoria sociologică iar altele sînt determinate de obiectul studiat.

In al doilea rînd, metoda poate primi un înțeles mai îngust, de modalitate efectivă de investigare a realității. Pornesc de la constatarea că științele naturii uzează de două metode de investigație empirică : observația și experimentul. Acestea, rămîn , la urma urmei, metodele de bază și ale sociologiei. Dar, avînd în vedere un specific cu implicații metodologice deosebite, și anume faptul că în științele socio-umane subiectul cercetător poate intra în relații de comunicare cu obiectul investigat, prin intermediul limbajului (comunicare ce poate lua diferite forme, de la o discuție față în față pînă la receptarea conținutului unor documente), observația ia unele forme ce se îndepărtează mult de ceea ce este în științele naturii, forme care, prin autonomizare , pot primi titlul de metode (specifice) ale stiințelor sociale.

La rîndul lor, metodele se pot subdiviza în funcție de tehnicile folosite, adică de operațiile concrete de înregistrare și manipulare a faptelor. In continuare, tehnica presupune, de cele mai multe ori, un instrument pentru a se putea realiza, cum ar fi, în cazul științelor sociale, chestionarul, ghidul de interviu, teste etc.

O clasificare a metodelor sociologiei, cu anumite subdiviziuni determinate de tehnicile folosite, ar putea fi următoarea:

a) Observația propriu-zisă

- exterioară (cînd cercetătorul nu se implică în viața comunității studiate);
- prin coparticipare (cînd se implică)

b) Interviul (dialogul față în față, urmărindu-se aprofundarea unor aspecte printr-o discuție relativ liberă)

c) Ancheta (caracteristica metodei este utilizarea chestionarului ca instrument de cercetare)

- directă (cînd chestionarul este completat de așa-numiții "operatori de anchetă" sau de persoane din echipa de cercetare)

- indirectă (autocompletarea chestionarelor)

d) Analiza documentelor

e) Experimentul

Termenul de procedee va fi folosit, de regulă, pentru a desemna operații și modalități de prelucrare a datelor unei cercetări empirice. Dar o utilizare riguroasă a lui în acest sens ridică numeroase probleme, dată fiind varietatea extrem de mare a mijloacelor de prelucrare a datelor, oferite mai ales de statistică. Astfel, termenul de procedeu ar trebui folosit pentru operații de anvergură foarte diferită, mergînd de la un simplu calcul de procent pînă la modele matematice sofisticate. De aceea - și pentru că există o destul de netă diferență între înregistrarea datelor pe teren și prelucrarea lor ulterioară - consider că nu se ridică nici o problemă dacă se reaiu termenii de "metodă" și "tehnică", mai ales cu funcții stilistice, pentru variația limbajului, și în acest din urmă domeniu, atunci cînd nu există riscul nici unei confuzii asupra semnificatului.

In fine, metodica cercetării este concepută - într-un sens - a fi strategia concretă întrebuințată în cursul unei cercetări efective, incluzînd metodele, tehnicile instrumentele utilizate precum și alte aspecte de ordin organizatoric. In același timp, metodica cercetării este o disciplină cu caracter normativ care, bazată pe experiența de cercetare empirică realizată pînă în prezent, încearcă să degaje anumite reguli, norme etc. a căror respectare este necesară pentru asigurarea unei bune desfășurări a cercetării și pentru realizarea unei eficiențe cognitive sporite. Ea prescrie, de exemplu, regulile pentru întocmirea unui chestionar, precizează condițiile concrete în care anumite metode sau tehnici cu șanse mai mari de a da rezultate pozitive, evidențiază capcanele metodice în care cercetătorii cu mai puțină experineță pot cădea etc. Dată fiind finalitatea aceasta forate concretă a metodicii, rezultă că și asertiunile sale au ca obiect niște aspecte foarte

concrete. Vreau să spun prin asta că metodica se preocupă nu atît de valenţele cognitive ale metodelor - nici în sensul general al termenului de metodă, nici în cel mai restrîns din care a rezultat enumerarea de mai sus - , ci în primul rînd de utilizarea efectivă a tehnicilor şi instrumentelor de cercetare sau, mai exact, spus, se preocupă de metode doar prin prisma modului în care acestea se concretizează în tehnici şi instrumente.

3.2. Valoarea reflexiei metodologice în sociologie

Este greu de găsit o altă ştiinţă despre al cărui domeniu metodologic să se fi scris mai mult decît despre cel al sociologiei. Situaţia aceasta a constituit nu o dată ţinta unor critici serioase sau a unor ironii, prin care sociologii erau îndemnaţi să producă mai multe rezultate pozitive despre realitatea socială şi nu numai discursuri nesfîrşite despre valoarea cutăriei metode sau tehnici de cercetare.

Abordarea domeniului metodologic al sociologiei s-a făcut şi se face, în general, din două unghiuri principale:

a) din cel al metodologiei şi

b) din cel al metodicii.

In primul caz, se analizează valoarea de cunoaştere a diferitelor metode şi tehnici; se discută probleme precum posibilitatea măsurării fenomenelor sociale, a explicaţiei cauzale, a experimentării etc. Din păcate, se amestecă aici probleme de natură foarte diferită, genul acesta de discurs fiind alimentat, în egală măsură, de consideraţii filosofice speculative şi de constatări pertinente desprinse din analiza riguroasă a cercetărilor efective, cum preconiza Lazarsfeld. In al doilea, se prezintă, cum spuneam, regulile dobîndite în practica cercetării concrete.

Cele două tipuri de abordare apar deseori îngemănate în una şi aceeaşi lucrare metodologică şi de aceea ele, în practică sînt destul de greu de separat. Insuşi cursul de faţă, deşi mă voi strădui, mai ales în secţiunea ce urmează, să mă axez cu preponderenţă pe problemele metodice ale cercetării, este un complex de chestiuni metodologice (de exemplu, capitolul despre experiment), metodice (cel despre anchetă) şi de probleme care aparţin ambelor

genuri dar şi unei discipline autonome cum este statistica (ultimele părţi ale cursului).

Proliferarea impresionantă a literaturii metodologice nu poate fi explicată numai prin cauze subiective, ţinînd de comoditatea cercetătorilor, care ar evita problemele dificile ale cercetării concrete şi s-ar refugia în cîmpul speculaţiilor metodologice, ci se datorează şi unor factori mai generali, rezultînd din specificul disciplinei. In acest sens, se pot aduce cel puţin două argumente esenţiale:

a) Mulţi sociologi au considerat şi consideră că reflexia metodolofică reprezintă pentru disciplina pe care o slujesc o şansă de evitare a unei serii de ocolişuri sau impasuri ce se ivesc în mod firesc pe drumul dezvoltării normale a unei ştiinţe. Avînd în faţă ştiinţe cu un statut bine consolidat, sociologia poate folosi experienţa acestora atît în domeniul construcţiei teoretice, cît şi în cel al constituirii şi utilizării arsenalului metodologic.

b) Specificul sociologiei şi al celorlalte ştiinţe sociale, discipline care au de a face cu elemente precum scopuri, valori, acţiuni etc. ce nu se întîlnesc în ştiinţele naturii, a implicat, cum era şi firesc să se întîmple, discuţii numeroase, care au făcut să crească volumul scrierilor metodologice cu privire la mijloacele cele mai eficace de cercetare. S-a urmărit , în mod cu totul deosebit, clarificarea următoarelor două aspecte:

- măsura în care metodele tradiţionale din ştiinţele naturii sînt utilizabile (sau pot fi făcute utilizabile) în cunoaşterea fenomenelor sociale;
- raportul dintre mijloacele de cunoaştere specifice ştiinţelor sociale şi cele din ştiinţele naturii.

Sînt probleme care au făcut să curgă multă cerneală şi care sînt şi astăzi, după mai bine de un secol, dezbătute cu aceeaşi pasiune de către filosofi şi sociologi. Cele două direcţii fundamentale, care s-au constituit şi care au divizat lumea sociologică, poate mai evident - cel puţin la suprafaţă - decît înseşi concepţiile teoretice, sînt, de regulă, numite monismul sociologic şi dualismul sociologic. Prima direcţie pleacă de la ideea unităţii metodelor de cercetare empirică indiferent de domeniul fenomenelor abordate, pe cînd a doua exacerbează distincţia dintre ştiinţele

naturii și cele socio-umane (sau, cu alți termeni, ale culturii, ale spiritului etc.) atît pe linia obiectului, cît și a metodelor. Este clar că aceste concepții despre metodă interferează cu cele filosofice ale autorilor, cu construcțiile teoretice asupra fenomenelor sociale, de unde rezultă o mare diversitate în cadrul fiecărei orientări fundamentale.

Scopul cursului de față fiind unul tehnico-metodic, nu vom ataca interesantul cîmp de probleme de acest gen; am ieși astfel chiar și din sfera metodologiei, trecînd în cea a epistemologiei și filosofiei științei. (Pentru detali, a se vedea, dintre lucrările mai recente, cea a lui L. Vlăsceanu , 1982). Subliniez doar , ceea ce s-a putut deja înțelege din cele spuse pînă acum. că opțiunea mea este pentru poziția monistă, prin urmare, pentru unitatea metodelor de cercetare empirică, cu singura precizare că acestea trebuie adecvate, în forma lor concretă de utilizare, la obiectele , fenomenele studiate și la raportul dintre obiect și subiect ce se instituie în procesul cunoașterii.

In măsura în care științele sociale aspiră la statutul de știință, ele trebuie să se supună rigorilor acestui statut. Una dintre cerințele elementare ale oricărei aserțiuni științifice despre un dat obiectiv constă în posibilitatea verificării acesteia de către orice alt cercetător. Or, o atare caracteristică devine posibilă numai în măsura în care se face apel la metode clare, standardizate de cunoaștere, așa cum utilizează de secole științele naturii. Esențialul, cred eu, în deosebirea dintre monismul și dualismul metodologic, nu constă în faptul că ultimul ar propune alte metode de cercetare în domeniul fenomenelor sociale, ci în acela că, odată cu repudierea metodelor naturaliste, este atacat, în esență, însuși spiritul științific al disciplinelor socio-umane.

Cap.II. ETAPELE CERCETARII SOCIOLOGICE

Aparent, există o mare diversitate de opinii în literatura metodologico-tehnică privind etapizarea unei cercetări sociologice empirice, numărul fazelor ce trebuie realizate pe parcursul întregii investigaţii variind, în opinia autorilor, de la 3 la 15 sau chiar mai multe. Spun că această diversitate este doar aparentă pentru că, de fapt, aproape de fiecare dată sînt precizate destul de exhaustiv momentele principale ale cercetării, numai că ele apar cu statute diferite: de etape propriu-zise sau subetape. In fond, o discuţie prelungită asupra acestei etapizări n-ar avea decît un caracter scolastic; important este, cum spuneam, ca momentele să fie enumerate şi analizate, iar structurarea nu are, la urma urmei, alt scop decît să faciliteze înţelegerea discursului.

Orice autor de manual va trebui să opteze pentru o anumită formulă, desigur nu cu totul întîmplător, ci în virtutea unei concepţii şi perspective proprii - nu neapărat originale - asupra modului de realizare a cercetării. In ceea ce mă priveşte, cred că o formulă cu un caracter de mai mare generalitate - care să surprindă filmul desfăşurării oricărei investigaţii sociologice şi nu numai a anchetelor -ar fi cea care porneşte iniţial de la trei mari etape: pregătirea cercetării, culegerea informaţiei şi valorificarea informaţiei. Acestea pot fi, la rîndul lor, subetapizate în funcţie de metodicele concrete folosite în cercetările sociologice, dar, chiar şi în cadrul lor, pot fi găsite elemente comune oricărei investigaţii.

Ultimele două etape sînt foarte strîns legate de metodele şi tehnicile efectiv utilizate în practica cercetării şi, prin urmare, amintitele elemente comune s-ar putea decela doar cu mare greutate; în schimb, etapa iniţială pare a se preta mai uşor la o analiză sistematică din punctul de vedere al structurării operaţiilor ce le incumbă. Din acest motiv, structura capitolului de faţă va fi deosebit de dezechilibrată, majoritatea comentariilor



referindu-se la faza pregătitoare a cercetării, mulţumindu-m
ca la sfîrşit să fac unele precizări legate de valorificarea in-
maţiei, omiţînd complet, în acest moment, consideraţiile despre cu-
legerea informaţiei, care vor apare în capitolele dedicate metode-
lor de cercetare sociologică.

1. <u>Pregătirea cercetării</u>

Cred că sînt puţine manuale şi tratate de metode şi teh-nici care să nu insiste de o manieră cu totul deosebită asupra importanţei pe care etapa pregătitoare o are asupra desfăşurării efective a cercetării şi asupra rezultatelor finale obţinute. Chiar şi fără asemenea insistenţe, oricine poate să-şi dea seama că rezultatul muncii sale depinde de gradul de perfecţionare a instrumentelor utilizate, de cunoştinţele pe care le are despre obiectul de studiat, de felul în care este pregătită şi organiza-tă din punct de vedere administrativ o investigaţie ce reuneşte adesea un număr mare de persoane etc. In ciuda tuturor avertismen-telor şi a evidenţei, întîlnim, nu odată, şi nu numai în cazul începătorilor în ale muncii sociologice, investigaţii pornite cu ochii închişi, cercetări pentru care se cheltuieşte un volum i-mens de muncă şi de mijloace materiale dar care ajung la rezulta-te minore tocmai din cauza neglijării pregătirii satisfăcătoare a lor. E vorba, de bună seamă, în cele mai multe cazuri, de anche-te pe bază de chestionar, a căror simplitate aparentă poate duce foarte uşor în eroare. <u>Carenţele de construcţie a chestionarului, de instruire a operatorilor</u>, de <u>alegere a eşantionului ş.a.m.d.</u> vor avea implicaţii negative asupra eficienţei anchetei, reducîn-d-o uneori aproape la zero.

Bineînţeles că pregătirea temeinică a unei cercetări constituie o condiţie fundamentală cu caracter de principiu pentru orice studiu indiferent de disciplina ştiinţifică la care aparţine. Dar, în acelaşi timp, pe lîngă remarcile de mai sus că există des-tule cercetări sociologice complet sau aproape complet ratate din lipsa unor preparative corespunzătoare, trebuie să înţelegem şi faptul că nu toate investigaţiile pot fi, obiectiv vorbind, la fel de bine pregătite. Deci gradul în care se realizează exigenţele fazei la care mă refer depinde de foarte mulţi factori şi nu numai



de sociolog, adică de priceperea sau nepriceperea lui, de precauţiunile pe care le ia sau nu, ori - de ce nu ? - de onestitatea sau lipsa lui de onestitate profesională. Vreau să spun că există cercetări în care o serie de condiţii independente de sociolog nu permit realizarea tuturor momentelor pregătitoare: lipsa unor studii similare, deci noutatea în abordarea fenomenului, care ne privează nu numai de posibilitatea comparării datelor obţinute dar şi de utilizarea unor concepte deja forjate în cercetarea empirică, de instrumente standardizate, de ipoteze cu o verosimilitate ridicată etc. poate constitui o cauză ce conduce la rezultate mai puţin consistente. Astfel de investigaţii sînt numite uneori studii cu caracter explorativ spre a le deosebi de cele denumite cu caracter explicativ. Probabil că eticheta folosită pentru acestea din urmă nu este prea fericit aleasă, deoarece oricît de modest ar fi un studiu, el nu poate renunţa - fără a-şi pierde statutul de cercetare ştiinţifică - la ideea explicării faptelor studiate. Oricum, distincţia de mai sus pune în evidenţă faptul că în anumite situaţii caracterul explicativ este mai slab sau mai puternic, în funcţie de posibilitatea realizării unei temeinice etape pregătitoare a cercetării.

Pe de altă parte, sociologul se poate lovi de obstacole de timp, de bani, etc. - mai ales în cazul cercetărilor contractuale - ceea ce-l conduce la realizarea unor investigaţii în pripă şi fără mijloace materiale şi umane corespunzătoare. Or, se poate afirma cu destul temei că atunci cînd sociologul este presat de timp sau de lipsuri de mijloace materiale - dar mai ales de timp - el va renunţa în primul rînd şi în cea mai mare măsură la o parte din activităţile reclamate de prima fază a cercetării, fiindcă lucrul acesta poate fi cel mai uşor de ascuns de ochii beneficiarului.

Să trecem la prezentarea principalelor operaţii ce se cer realizate în etapa pregătitoare, acordînd un spaţiu mai important celor în care aspectele metodice sînt mai proeminente.

1.1. Alegerea problemei (temei) de cercetare

Este normal că ne aflăm aici în punctul inițial al oricărei investigații. Din perspectiva cursului de față, chestiunea esențială ce trebuie subliniată rezidă în necesitatea circumscrierii corecte și cît mai exacte a problemei de cercetat. Orice imperfecțiune a acestei operații are consecințe negative atît asupra muncii de cercetare propriu-zisă, cît și, evident, asupra rezultatelor studiului.

Din alt punct de vedere, aș dori să menționez că deși alegerea temei de cercetare este în aparență un act ce depinde numai de liberul arbitru al cercetătorului, în realitate se dovedește a finunul ce i se impune acestuia din urmă într-o măsură suficient de mare. Este normal ca în alegerea respectivă să intervină și subiectivitatea cercetătorului, preferințele, experiența și înclinațiile sale, dar, de obicei - pentru că sînt foarte rare cazurile cînd cercetătorii sînt cu totul independenți de anumite instituții sau organisme sociale - intervin cel puțin patru elemente constrîngătoare în alegerea temei de studiu:

(i) Tematica de cercetare a instituției în care sociologul lucrează trebuie să se orienteze spre probleme de interes major pentru societate. Această tematică este supusă, de regulă, aprobării forurilor superioare instituției respective, foruri care ghidează și orientează cercetarea în domeniul în cauză.

(ii) Cercetătorul dispune totdeauna de un anumit volum de mijloace materiale și este evident că el nu va putea aborda decît acele teme a căror investigare este posibilă în cadrul plafonului de mijloace pe care le are la dispoziție.

(iii) In al treilea rînd, aproape orice cercetare sociologică este o muncă de echipă. Este normal ca în funcție de numărul cercetătorilor, de profilul preocupărilor și specializării lor, conducătorul echipei va fi influențat să opteze pentru anumite teme. Iar din punctul de vedere al membrilor echipei, este de asemenea clar că fiecare dintre ei va fi într-o anumită măsură marcat de profilul colectivului din care face parte.

(iv) Un alt set de factori care pot influenţa alegerea problemei de cercetare este constituit din elemente precum: gradul de dezvoltare a diverselor ramuri ale sociologiei, tradiţiile existente, apariţia pe "piaţa sociologică" a unor teme la modă, bibliografia şi datele statistice de care se dispune etc.

In cazul cînd sociologul lucrează în cadrul unei instituţii neacademice, atunci cu siguranţă că întreaga sa activitate de cercetare empirică va fi determinată de profilul instituţiei respective, deci domeniul de alegere se reduce simţitor. Sociologul de întreprindere va fi chemat să abordeze doar probleme care prezintă interes pentru acea intreprindere, preocupîndu-se de teme de sociologia muncii şi sociologia industrială; cel încadrat la un centru de proiectare în domeniul sistematizării se va ocupa în mod precumpănitor de teme de sociologie urban-rurală, demografie etc.

1.2. Studiul bibliografiei

Probabil că pentru un fizician sau alt cercetător din ştiinţele mature o asemenea recomandare de a studia bibliografia înainte de a se apuca de o cercetare ar părea cel puţin superfluă, dacă nu de-a dreptul ridicolă. Din păcate, în cadrul sociologiei ea trebuie făcută mereu şi chiar subliniată din cînd în cînd şi asta nu doar pentru faptul, clar şi îndeobşte cunoscut, că nu te poţi angaja într-un studiu serios fără a cunoaşte ceea ce au făcut şi alţii în domeniul respectiv. Situaţia sociologiei este de aşa natură încît permite saltul peste această etapă (sau, mai corect spus, lasă iluzia că s-ar putea trece peste ea). Cunoscînd precaritatea multor cercetări, sociologul poate exagera uneori în sensul de a considera că tot ceea ce au făcut alţii este lipsit de valoare şi deci se poate dispensa de experienţa lor. O altă faţă a aceleiaşi chestiuni se poate manifesta chiar atunci cînd sociologii cunosc rezultatele confraţilor lor, dar în cercetările proprii, dacă nu le ignoră, încearcă să le "depăşească" printr-un exces de originalitate. Probabil că avem de a face aici cu o influenţă nefastă a pregătirii în general filosofice a sociologilor noştri, influenţă concretizată în iluzia că originalitatea se confundă cu valoarea. Din nefericire - sau fericire - în ştiinţă lucrurile

nu stau deloc aşa, pentru că aici aspectele cu totul originale sînt mai degrabă excepţii; regula o constituie munca asiduă de culegere şi interpretare a informaţiei în cadre bine constituite, muncă ce nu se poate realiza decît printr-o cunoaştere exactă şi clară a ceea ce s-a realizat, printr-o continuitate a preocupărilor şi metodelor de lucru.

Prin urmare, studiul bibliografic permite sociologului o ameliorare a rezultatelor muncii sale din perspectiva mai multor factori, dintre care amintesc:

a) delimitarea precisă a obiectului cercetării;

b) utilizarea unor teorii sau scheme explicative a căror validitate a fost testată;

c) folosirea unor ipoteze de lucru rezultate dintr-o cunoaştere prealabilă a domeniului;

d) aplicarea unor metode, tehnici, instrumente şi procedee care au condus deja la rezultate consistente în domeniul respectiv.

Una dintre consecinţele de cea mai mare valoare pentru dezvoltarea sociologiei pe care o are studierea aceluiaşi fenomen cu metode, concepte şi cadre teoretice similare o constituie comparabilitatea rezultatelor studiilor sociologice empirice. Este realmente o problemă de o deosebită însemnătate, fără rezolvarea căreia este imposibil să se îndeplinească condiţia de bază a dezvoltării unei ştiinţe; cumulativitatea cunoştinţelor dobîndite. Din păcate, sociologii păcătuiesc, cum spuneam, de excesul de originalitate, au mania "delimitărilor" permanente de preocupările altora şi asta nu numai în cîmpul cercetării teoretice dar şi al celei empirice. De aici munţii de date empirice culese de aproape trei sferturi de secol de sociologie de teren regulată, din care însă nu au apărut multe concluzii de mare amplitudine.

1.3. Precizarea conceptelor

Este o operaţie strîns legată de precedenta, căci ea nu se poate realiza fără cunoaşterea încercărilor - reuşite sau nereuşite - ale predecesorilor. Despre importanţa conceptelor, în general într-o cercetare ştiinţifică , s-a vorbit atît de mult încît

nu are rost să mai insist aici asupra ei. Menţionez doar că, de regulă, conceptului i se atribuie trei funcţii importante pentru cercetarea empirică:

- să organizeze percepţia
- să ghideze observaţia şi
- să prezică fapte noi.

Aşa cum am mai spus-o, în sociologie avem de a face o penurie de concepte bine precizate. Există o mare ambiguitate a termenilor, împrumutaţi în imensa lor majoritate din limbajul comun. Din această cauză, utilizarea lor fără o definire prealabilă devine, dacă nu imposibilă, în orice caz ineficientă în munca de cercetare. Desigur că nici definirea de fiecare dată în mod original a conceptelor nu este o modalitate prea elegentă şi economicoasă pentru dezvoltarea unei ştiinţe.

Dacă s-ar putea măsura gradul de dezvoltare al ştiinţelor nu cred că ar fi greu să se pună în evidenţă o corelaţie negativă puternică între acest indicator şi numărul conceptelor folosite de diferite ştiinţe. Iar în sociologie, trebuie s-o recunoaştem, există o inflaţie de noţiuni. De pildă, ca să reiau un exemplu vehiculat în literatura metodologică, în 1931 un autor american, pe nume Eubank, a întocmit o listă cu noţiunile principale folosite în 8 lucrări de sociologie cu caracter general, ale unor autori diferiţi, listă care cuprindea în total 146 de concepte. Dintre acestea nici unul nu a putut fi regăsit în toate cele 8 lucrări; în 7 lucrări s-a regăsit unul singur comun, în 6 lucrări - două iar în 5 lucrări - opt. Din totalul de 146 noţiuni, 83 erau specifice cîte unei singure lucrări. Si nu sînt convins că situaţia s-a ameliorat radical de atunci şi pînă azi. Chiar şi în cazul cînd unele noţiuni au o circulaţie mai largă, sensul şi semnificaţia lor variază sensibil de la autor la autor.

Se impune deci cu necesitate faptul ca la începutul oricărei cercetări sociologice de teren să se definească principalele concepte cu care se lucrează, iar dacă este posibil această definire să se facă pe linia uniformizării limbajului, deci în mod similar altor cercetări, pentru comparativitatea datelor. Este necesară definirea precisă la început deoarece de semnificaţia acordată conceptelor depinde în mod hotărîtor întregul proces de culegere, sistematizare, prelucrare şi analizare a informaţiei.

1.4. Operaţionalizarea conceptelor

Pentru orice ştiinţă empirică, una dintre problemele sale fundamentale este aceea de a stabili modalităţile prin care se realizează legătura dintre universul teoretic şi realitatea concretă, să-i zicem, reflectată. In cadrul acestei chestiuni generale, un loc central îl ocupă problema operaţionalizării conceptelor, ca ansamblu de operaţii prin care însuşirile definitorii ale noţiunilor pot fi identificate sau măsurate în universul empiric. O abordare teoretică de profunzime a acestei probleme ne-ar duce foarte departe la aspecte de epistemologie, filosofia şi logica ştiinţei, metodologie generală etc.
In scopul înţelegerii activităţii practice de operaţionalizare, voi schiţa doar cîteva repere.

Mai întîi - şi iarăşi fără a intra în chestiuni filosofice adînci, în speţă de teoria definiţiei - se cuvine observat că în sociologie apar cu cea mai mare frecvenţă două tipuri de noţiuni: unele, pe care să le numim de tip atributiv a) şi altele, de tip variabil. b) Primele sînt cele studiate pe larg de logica clasică bivalentă, prin punerea în evidenţă a sferei şi continutului noţiunii. Ele presupun, în fond, prezenţa unui univers de indivizi empirici care poate fi decupat în două submulţimi: cea formată din "obiectele" ce poartă însuşirea (sau însuşirile) cerută de conţinutul noţiunii - şi formînd sfera acesteia - şi cea formată din cele care nu posedă trăsătura respectivă. Aşa de exemplu, în cercetările noastre avem de a face cu noţiuni ca "muncitor" (membru al clasei muncitoare), "activ", "retribuit", "ateu", "căsătorit" ş.a.m.d. toate acestea fiind evident de tip atributiv.

b) Noţiunile de tip variabil presupun, de asemenea, un univers de "obiecte" empirice (care, ca şi în cazul precedent, pot fi indivizi umani, grupuri sociale, instituţii, comunităţi sau relaţii umane, acţiuni sociale etc.) dotate cu o anumită trăsătură ce nu se mai evidenţiază doar prin cele două alternative - prezenţă/absenţă -, ci printr-o variaţia graduală în intensitate. Aşa sînt conceptele de satisfacţie în muncă, de integrare socială, de coeziune a unei colectivităţi şi multe, foarte multe altele. Chiar dacă unele noţiuni de tip variabil pot fi transformate, prin

simplificare, în altele de tip atributiv (vorbim despre indivizi satisfăcuţi de muncă sau nesatisfăcuţi, despre integraţi sau neintegraţi etc.), diferenţa este totuşi esenţială, mai ales în problema ce ne interesează aici.

Revenind acum tocmai la această problemă, vom mai observa că literatura metodologică face distincţia dintre aşa-numitele concepte <u>ideale</u>, <u>abstracte</u>, <u>latente</u>, <u>neoperaţionale</u> etc. "care nu au referinţe observaţionale (empirice) nemijlocite" şi altele <u>empiric referenţiale</u>, <u>manifeste</u>, <u>concrete</u>, <u>operaţionale</u> etc. , ce "definesc entităţi direct observabile" (Vezi L.Vlăsceanu, 1982, pp.258-259). O astfel de delimitare, deşi rigidă, ca orice dihotomie, are meritul de a sugera foarte clar problematica operaţionalizării conceptelor. Intr-adevăr, un concept de tip variabil ca "vîrsta" poate fi imediat perceput în realitatea obiectivă (fie întrebînd individul în cauză sau persoane apropiate de acesta, fie consultînd certificatul de naştere sau alt document potrivit). Dimpotrivă, satisfacţia muncii sau calitatea vieţii unei persoane sînt practic nedecelabile printr-o operaţie empirică atît de simplă. Cum se procedează în asemenea cazuri ?

Dacă ne-am apleca asupra unor noţiuni obişnuite, utilizate în mod curent în limbajul comun, vom observa că, de fapt, fiecare dintre noi realizează operaţionalizări fără să-şi dea seama atunci cînd distinge, de exemplu, un bărbat de o femeie, o relaţie de rudenie de una de serviciu, un profesor de un elev, o stare de tristeţe de una de veselie etc. De multe ori oamenii nu au cunoştinţe prea clare despre conţinutul unor asemenea noţiuni, ceea ce nu-i împiedică să le folosească totuşi în mod curent şi cu relativ succes (adică fără să intre prea des în situaţii dilematice). Este probabil foarte greu pentru un om obişnuit să definească în mod acceptabil, de pildă, noţiunea de tristeţe, dar nimeni nu se îndoieşte de capacitatea sa de a diagnostica tristeţea (cel puţin în rîndul persoanelor ce-i sînt apropiate) prin semne care i se par absolut semnificative în comportamentul semenilor săi. Mai mult, chiar dacă e uşor de definit o noţiune, în practica zilnică nu este necesar - şi nici comod - să se folosească în operaţiile de identificare tocmai trăsăturile definitorii. O serie de alte aspecte, mai uşor de sesizat, sînt suficiente pentru a

se ajunge la acelaşi rezultat. Astfel, felul de a se îmbrăca, vocea, coafura, forma corpului etc. sînt elemente care în cvasitotalitatea cazurilor permit să se facă distincţia între o femeie şi un bărbat, cel puţin de la o anumită vîrstă în sus.

Conceptele mai complexe, care urmăresc aspecte tot atît de reale ca şi cele din exemplele amintite aici, se definesc, de regulă, prin trăsături mai generale, fiind rezultatul unor prelucrări şi rafinări ale altor concepte mai simple. In astfel de cazuri, apropierea lor de realitate nu se poate face decît prin constituirea unor trăsături simple care să poată fi identificate direct în empiric. Este vorba de un set de însuşiri, pentru că e aproape imposibil de găsit o trăsătură simplă care să fie echivalentă cu cea complexă din definiţie; numai prin "însumarea" mai multor aspecte parţiale se poate spera să se ajungă la scopul urmărit. Mai mult, gradul înalt de complexitate a unor concepte se datorează şi particularităţii lor de a acoperi dimensiuni distincte ale realităţii. Astfel, dacă vorbim de integrarea studenţilor din anul I în viaţa universitară, trebuie să avem în vedere cel puţin următoarele aspecte: integrarea în colectivul anului (grupei_ de studiu, adaptarea la stilul şi cerinţele activităţii profesionale universitare, integrarea în viaţa organizaţiei ASC, adaptarea la viaţa de cămin (dacă e cazul), adaptarea la stilul de viaţă al oraşului (cel puţin pentru cei nou veniţi în oraş) etc. Este evident că aceste paliere apar efectiv ca dimensiuni distincte, deoarece ele sînt practic independente unul de altul (un student poate fi considerat perfect integrat în raport cu unul şi total neintegrat în raport cu altul) şi de asemenea, este normal ca în încercarea de evaluare sintetică să se ţină seama de toate aceste dimensiuni. Pentru cercetarea concretă este nevoie ca fiecare dimensiune aleasă să fie, la rîndul său, operaţionalizată, adică transpusă în indici empirici.

Dar să-l urmărim mai bine pe P.Lazarsfeld, care, cu claritatea-i recunoscută, sintetizează etapele ce se cer parcurse pe drumul ce duce de la concept la studiul empiric (1971, pp.20-23), adică tocmai esenţa operaţionalizării.

(i) Iniţial avem o imagine sau o reprezentare destul de vagă a ceea ce înţelegem printr-o noţiune, imagine rezultată fie din decelarea unor trăsături comune subiacente mai multor

fenomene dispersate, fie din încercarea de a explica anumite constante.

(ii) A doua etapă constă în specificarea conceptului. Ea consistă în descompunerea reprezentării în componente, ce constituie diverse aspecte sau dimensiuni ale conceptului. Astfel de componente se deduc uneori din analiza teoretică a conceptului iar alteori rezultă din analiza empirică a corelaţiilor observate.

(iii) A treia etapă se referă la alegerea indicatorilor. Indicatorul nu este altceva decît acea însuşire simplă care poate fi detectată direct în realitatea empirică. De exemplu, dacă observăm că un om frecventează regulat biserica, acesta va fi un indicator destul de sigur al religiozităţii persoanei respective. Indicatorii pot fi, evident, de genuri foarte diferite, fie în funcţie de modalitatea de conectare cu realitatea, fie după tipul de informaţie pe care ei o furnizează (de la specificări de genul da/nu pînă la valori numerice, rezultat al măsurării).

(iv) A patra etapă este cea a formării indicilor. Indicii empirici, cu ungrad mai mic sau mai mare de generalitate, urmăresc să suprindă sub o formă numerică concentrată informaţia conţinută într-un set de indicatori. Dacă o dimensiune oarecare este operaţionalizată printr-un singur indicator ce permite măsurarea, atunci se obţine direct un indice pentru aspectul în cauză (de pildă, media valorilor individuale). Dacă sînt mai mulţi indicatori, atunci sau se traduce fiecare într-un indice şi aceştia din urmă se "condensează" apoi în unul singur (printr-o procedură de medie, de exemplu) sau se combină indicatorii între ei astfel încît să se ajungă la o singură valoare finală - indicele dimensiunii. In sfîrşit, ultima etapă de imaginat ar fi sintetizarea indicilor dimensiunilor în unul singur, general, capabil să reflecte întreaga problematică a conceptului. S-ar ajunge astfel la indici de calitate a vieţii, de satisfacţie, de integrare etc.

Desigur că tehnologiile de obţinere a unor asemenea indici sintetici sînt destul de complicate, problemele metodologice şi tehnice ridicate de operaţia respectivă nu sînt uşor de rezolvat. Există o literatură destul de întinsă legată de problematica indicatorilor şi indicilor, dar în cursul de faţă nu putem zăbovi

mai mult asupra ei, mai ales acum în partea de debut. Astfel de chestiuni ar putea fi cu mai mult folos abordate după introducerea metodologică şi statistică ce constituie, în fapt, obiectul întregului nostru curs. Ceea ce se poate face aici cu destul folos este să se realizeze o clarificare a lucrurilor prin exemple concrete.

Intr-o cercetare de sociologie industrială, C.Zamfir şi colaboratorii (1980) au urmărit, printre altele, operaţionalizarea noţiunii: "satisfacţia muncii". Sistemul de indicatori folosit în acest scop a cuprins o baterie de 8 întrebări din chestionarul folosit, întrebări care cereau subiecţilor să exprime gradul de satisfacţie încercată în legătură cu următoarele aspecte:

- profesiunea,
- intreprinderea,
- componenţa grupului de muncă,
- şeful direct,
- conţinutul muncii,
- promovarea efectivă (pînă în prezent),
- posibilităţile de promovare (în viitor),
- retribuţia

Intrebările erau de genul: " In general, cît de satisfăcut sînteţi de persoanele din grupul dv.?", cu răspunsuri mergînd pe o scală cu 7 trepte, de la foarte puţin la foarte mult, fiecare treaptă primind o valoare numerică de la 1 la 7. Din cumularea valorilor atribuite de fiecare individ la cele 8 întrebări, s-a obţinut o cifră totală cu rolul de indice global de satisfacţie.

După prelucrarea datelor şi efectuarea unor analize statistice, s-a ajuns la concluzia că indicatorii folosiţi tind să circumscrie trei dimensiuni ale conceptului:

a) satisfacţie în raport cu relaţiile din colectivul de muncă;

b) satisfacţia cu conţinutul şi cadrul de desfăşurare a muncii;

c) satisfacţia cu beneficiile personale asociate muncii.

Este evident că o asemenea procedură de construcţie a dimensiunilor unui concept este total dependentă de natura şi numărul indicatorilor aleşi (în cazul de faţă, întrebările chestionarului). Rezultatul operaţiei va fi variabil, de la cercetător la cercetător.

cetător, neexistînd garanţia depistării "adevăratelor" dimensiuni ale conceptului . Dar nici procedeul invers - de stabilire a priori a dimensiunilor - nu este scutit de pericolul subiectivismului. In absenţa unei teorii clare, o analiză, oricît de atentă ar fi ea, nu poate garanţa identificarea tuturor dimensiunilor unui concept. Eventuala consolare că o analiză ulterioară a indicatorilor aleşi va confirma sau infirma dimensionarea apriorică (deci ne va ajuta în aprecierea validităţii sale) este pur şi simplu o iluzie. Dacă dimensiunile sînt dinainte date, numai un cercetător lipsit de o minimă experienţă (ca să nu zic de inteligenţă) va alege indicatori ale căror corelaţii să fie mai slabe între perechile formate din indicatorii ce intră într-o aceeaşi dimensiune, decît între perechile aparţinînd la dimensiuni diferite. (In general, identificarea dimensiunilor a posteriori sau validarea celor introduse a priori se face cu ajutorul unor proceduri bazate pe coeficienţii de corelaţie, coeficienţi care trebuie să aibă valori mai ridicate între variabilele cuprinse în aceeaşi dimensiune decît între cele provenite din dimensiuni distincte).

Oricum s-ar proceda, este limpede că un concept odată operaţionalizat, odată specificate şi urmărite anumite dimensiuni ale sale, rezultatele cercetării empirice vor descrie numai aspectele reţinute de dimensiunile şi indicatorii utilizaţi. De aici rezultă cu claritate că extrapolarea constatărilor cercetării la întreaga semnificaţie a conceptului este ilicită şi că, prin urmare, de cele mai multe ori, investigaţia empirică - tocmai prin surprinderea unui cîmp de dimensiuni - sărăceşte, simplifică realitatea. Esenţial aici este de a găsi o formulă simplificatoare acceptabilă, în sensul că vor fi totuşi surprinse elemente esenţiale şi că ceea ce se pierde printr-o astfel de sărăcire este înzecit compensat de cîştigurile obţinute prin precizia măsurătorilor, profunzimea analizelor etc.

Aşezarea problemei operaţionalizării conceptelor în rîndul elementelor de bază în pregătirea cercetării şi sublinierea importanţei acestei probleme pentru eficienţa investigaţiei nu înseamnă acceptarea poziţiei filosofice cunoscută sub numele de operaţionalism. Această concepţie - o variantă a neopozitivismului - susţine, în esenţă, ideea că definirea noţiunilor se reduce la specificarea operaţiilor efectuate în universul datelor empirice. Caracterul său

relativist şi subiectiv a fost pus cu claritate în evidenţă de filosofia marxistă, dar, ca şi întreg pozitivismul, operaţionalismul a avut şi el o influenţă semnificativă asupra perfecţionării metodelor de cunoaştere în ştiinţele socio-umane.

1.5. Stabilirea ipotezelor de lucru

Ne găsim şi aici în faţa unei probleme ce se află la interferenţa epistemologiei, filosofiei ştiinţei şi metodologiei. In toate aceste domenii s-a abordat pe larg conţinutul conceptului de ipoteză, precum şi rolul şi locul acesteia în cunoaşterea ştiinţifică. Ceea ce doresc să menţionez aici este o idee ce-şi găseşte loc adesea în lucrările metodologice. O investigaţie ştiinţifică nu se realizează niciodată în afara sau în lipsa unor ipoteze. Este o iluzie faptul că ipotezele ar lipsi din cercetare; ele există sub o formă inconştientă sau nespecificată chiar atunci cînd cercetătorul se străduieşte să studieze o realitate în mod cît mai "obiectiv", fără "idei preconcepute". De exemplu, construcţia oricărui chestionar se bazează pe o mulţime de ipoteze. Faptul că se pune o întrebare şi nu alta înseamnă că cercetătorul a ales ca probabilă o legătură între fenomenele X şi Y şi nu între X şi Z. Faptul, mai general, că se adoptă anumite metode şi tehnici de cercetare se bazează tot pe ipoteze legate de specificul realităţii studiate ş.a.m.d. Deci, practic, orice operaţie, oricît de măruntă ar fi, este pînă la urmă impregnată, susţinută şi ghidată de anumite ipoteze.

Concluzia care se degajă este evidentă. Decît să se lucreze cu asemenea ipoteze neconştientizate, deci necontrolate, este preferabil ca înaintea oricărei cercetări să se facă un inventar cît mai complet al ipotezelor implicate. Ar mai fi de adăugat la aceasta că ipotezele nu apar numai înaintea muncii empirice, deci nu numai în legătură cu datele culese, ci şi după etapa obţinerii informaţiei, atunci cînd se face interpretarea datelor. De pildă, dacă în etapa iniţială presupunem existenţa unei corelaţii între două fenomene, în final, dacă respectiva corelaţie există, se vor emite ipoteze legate de raportul mai adînc de legătură între fenomene. Astfel de ipoteze pot fi testate empiric în

cadrul cercetării sau pot rămîne deschise şi deci apar ca rezultat al unei investigaţii.

1.6. Alegerea metodelor şi tehnicilor

Reprezintă, desigur, un alt moment important al cercetării, care se realizează practic simultan cu alegerea problemei de cercetat, deoarece fiecare obiect antrenează tehnici de investigaţie specifice sau, mai bine spus, metodele şi tehnicile se cer particularizate şi concretizate în funcţie de tema de cercetare. Am folosit pluralul în subtilul de mai sus deoarece adeseori în investigaţiile sociologice se recurge la o îmbinare de metode şi tehnici Metoda anchetei, de exemplu, poate fi realizată în variante tehnice diferite şi, în plus, ea poate fi şi este adeseori însoţită de o analiză a documentelor, de observaţie sau interviu.

In afara temei de cercetare, arsenalul metodico-tehnic utilizat se mai alege desigur şi în funcţie de concepţia teoretico-metodologică a cercetătorului, de experienţa sa, adică de felul în care se simte capabil să mînuiască diferite, metode, tehnici şi instrumente de cercetare. De asemenea, nu trebuie neglijată nici influenţa mediului sociologic general, care, în diferite momente istorice, impune recurgerea la metodici specifice de cercetare. In fine, nu se pot omite într-o asemenea enumerare nici mijloacele materiale şi umane precum şi timpul avut la dispoziţie de cercetător. Aceste mijloace externe constrîngătoare îşi fac simţită influenţa, după cum se vede, în aproape toate elementele ce constituie osatura desfăşurării investigaţiei sociologice. Oricît de mari ambiţii ar avea un cercetător, dacă nu are la dispoziţie bani sau personal calificat, nu va putea recurge la tehnologii scumpe sau care cer o pricepere deosebită în culegerea datelor. Probabil că acesta este şi unul dintre motivele care au impus într-o vreme ancheta ca metodă privilegiată în studierea fenomenelor sociale.

1.7. Construcţia instrumentelor

Este evident, un aspect ce decurge logic din cel anterior, instrumentul fiind acela care concretizează, materializează o tehnică oarecare. Prea multe lucruri generale de natură metodică nu se pot spune nici la acest punct, deoarece fiecare instrument de culegere a datelor ridică probleme specifice de întocmire. Despre instrumentul cel mai utilizat în sociologie - chestionarul - o să vorbim pe larg în capitolul destinat anchetei, întrucît o bună parte din problematica respectivei metode este legată de întocmirea, aplicarea şi prelucrarea chestionarului.

Voi aminti aici doar cele două cerinţe, fundamentale pe care trebuie să le îndeplinească, de fapt, orice instrument, indiferent de tipul de investigaţie empirică: fidelitatea şi validitatea. Pe scurt, vom înţelege prin fidelitatea unui instrument capacitatea sa de a reproduce în mod constant anumite însuşiri constante ale fenomenului. Validitatea este proprietatea instrumentului de a dezvălui cercetătorului ceea ce acesta urmăreşte să obţină în cercetare. Această din urmă cerinţă este mai "tare" decît prima, în sensul că un instrument poate să fie fidel fără a fi valid. De exemplu un test poate să dea rezultate identice aplicat la grupuri identice, dar să nu măsoare ceea ce-şi propune.

În speţă, referindu-ne la validitatea unui chestionar, se înţelege că, pentru realizarea ei, aspectul realităţii abordate trebuie surprins cît mai exact şi mai complet posibil în seturile de întrebări, adică acestea trebuie să fie relevante pentru problema investigată, iar mijloacele de înregistrare - scalele, de exemplu - să suprindă de o manieră cît mai precisă informaţia culeasă. Această adecvare a instrumentului la obiectul ştiinţific şi garantează o anumită formă de validitate, numită, de obicei, validitate externă. Se poate vorbi, în egală măsură, şi de una internă, ce derivă din construcţia instrumentului, împlicînd, în principal, fidelitatea acestuia: informaţiile obţinute să poată fi univoc clasificate şi, dacă e vorba de chestionare, teste etc., termenii folosiţi să fie univoc înţeleşi de către fiecare operator sau subiect.

Oricum, problema validităţii instrumentelor în cercetarea sociologică rămîne una dintre chestiunile de cea mai mare importanţă.

deoarece instrumentul nu numai că reprezintă o sursă principală de erori, dar, de modul în care fenomenele sînt traduse în indicatori, de calitatea scalelor prin care se evaluează însuşirile şi de alte calităţi de construcţie a instrumentului depinde calitatea rezultatelor, consistenţa acestora, în fond calitatea cunoştinţelor dobîndite. Cercetările metodologice mai recente au pus în evidenţă unele modalităţi de testare a fidelităţii şi validităţii instrumentelor sociologice. Din păcate, ele au încă o eficienţă scăzută şi, chiar şi aşa, putem spune că apelul făcut la ele de către sociologii noştri nu este totdeauna satisfăcător, pentru că nu rareori aceste proceduri sînt insuficient înţelese. Intîlnim destul de des formulări de genul: "consistenţa internă a chestionarului a fost demonstrată de..." sau "instrumentul posedă un înalt grad de validitate" etc., deci aprecieri foarte categorice apărute în urma unor operaţii formale ce reflectă puţin din fondul chestiunii, formulări care trebuie mereu privite cu rezervă şi, aş zice, cu atît mai multă rezervă cu cît calităţile instrumentului sînt mai puternic apreciate.

1.8. Delimitarea universului de studiu şi a eşantionului

E vorba aici, mai întîi, de a decupa din realitatea obiectivă totalitatea obiectelor purtătoare de însuşiri, obiecte ce constituie universul material al cercetării, aria, zona asupra căreia se vor extinde rezultatele investigaţiei. Operaţia este legată foarte strîns de problema definirii conceptelor şi de cea a construcţiei indicatorilor empirici.

Delimitarea riguroasă şi eficientă pentru cercetare a populaţiei sau universului de unităţi de studiat este o operaţie extrem de dificilă, ridicînd multe probleme de ordin practic şi teoretic pentru investigatorul realităţilor concrete. Intr-adevăr, acesta nu se poate mulţumi cu formulări generale sau abstracte, care pot oferi, în principiu, o imagine asupra cîmpului de investigat, dar care atunci cînd sînt puse efectiv la lucru se dovedesc incapabile să ofere criterii exacte de delimitare.

Să dăm un singur exemplu, aproape banal, dar care va oferi o imagine destul de concludentă asupra dificultăţilor cu car

se confruntă sociologul practician. I se comandă, să zicem, de către un studio de radio regional să testeze opiniile locuitorilor unui oraş despre programele postului de radio respectiv. Care va fi populaţia ce formează obiectul (material) al studiului ? Desigur că după o prealabilă definire foarte clară a noţiunii de "locuitor al oraşului X" va trebui să se elimine din populaţia astfel obţinută toţi locuitorii sub o anumită vîrstă (n-are sens să se pună problema pentru sugari). Pe de altă parte, opinia despre postul de radio în cauză nu poate fi cerută decît celor care, în general ascultă radioul şi, în particular, celor care acordă o atenţie relativ regulată emisiunilor postului nostru. Dar cum s-ar putea determina, de exemplu, într-un oraş de talia Clujului, ca să nu vorbesc de altele mai mari, cît de cît exact populaţia care ascultă programele de radio. Ne va duce gîndul, desigur, la faptul că, pentru a asculta, omul trebuie să fie ori în posesia unui aparat de radio, ori să aibă acces la el. Cît de greu este să se depisteze chiar şi numai posesorii aparatelor de radio oricine îşi poate imagina; evidenţele aflate la oficiile poştale sînt destul de incomplete: există indivizi care nu declară posesiunea receptorului, aşa după cum există alţii în evidenţă - desigur mai puţini - cu aparate de radio stricate sau înstrăinate de o anumită vreme, fiind deci în imposibilitatea de a se pronunţa asupra emisiunilor din ultima perioadă, chiar dacă ar fi fost ascultători ai postului respectiv. Problema accesului efectiv la radioreceptoare este şi mai greu de rezolvat căci într-o familie poate exista un radio dar unii dintre membri să nu asculte, tot aşa cum alte persoane care nu au în locuinţă receptoare să asculte în altă parte (la locul de muncă, la prieteni etc.). Practic, ar fi nevoie deci, numai pentru delimitarea populaţiei de radioascultători, fără specificarea posturilor frecventate, săse facă o investigaţie prealabilă în oraşul X. Nu este deci de mirare că de multe ori, în cazuri ca cel invocat mai sus, universul cercetării se conturează efectiv în cursul acesteia, lucrul posibil atunci cînd se porneşte de la un univers mai larg (incluzîndu-l pe al nostru), univers ce oferă indicatori de delimitare foarte precişi.

Eşantionarea se foloseşte în situaţiile în care populaţia este extinsă şi nu este nici posibil, nici eficient să se studieze

în totalitatea sa. O subpopulaţie aleasă cu grijă, pe baza unor principii statistice derivate din teoria matematică a probabilităţilor sau pe temeiul unor cunoştinţe dobîndite în practică, poate suplini în bune condiţii cercetarea universului întreg. Asupra acestei chestiuni, deosebit de importantă pentru investigaţia sociologică, voi reveni mai pe larg în ultima parte a cursului, arătînd atît modalităţile practice de alegere a eşantionului cît şi semnificaţia rezultatelor obţinute în cercetările selective.

1.9. Organizarea acţiunii practice de cercetare

Intrucît cercetările sociologice se efectuează cel mai adesea cu un personal destul de numeros şi pot avea întinderi temporale şi extinderi teritoriale destul de însemnate (de unde şi denumirea, folosită uneori, de "campanii" sociologice), este clar că pentru buna desfăşurare a acţiunii trebuie să se depună un efort organizatoric deosebit. Numai cine a organizat efectiv o campanie sociologică cu cel puţin 2o-3o de oameni în echipă poate aprecia foarte exact dificultăţile ce se cer depăşite în asemenea situaţii. Aceasta nu înseamnă că ele nu pot fi totuşi sugerate, ceea ce voi încerca să fac, de o manieră foarte succintă, în cele ce urmează.

Mai întîi, se impune o riguroasă planificare în timp a desfăşurării studiului, în care să apară foarte clar corelate toate etapele şi operaţiile ce se vor realiza. De asemenea, trebuie să se prevadă consecinţele întîrzierilor în finalizarea unor etape, ce se găsesc într-o măsură mai mică sub controlul cercetătorului, şi să se aibă în vedere soluţii de schimb.

In al doilea rînd, apare ca o necesitate stringentă planificarea modului de gospodărire a mijloacelor materiale de care se dispune: cît poate fi cheltuit pentru documentare, pentru multiplicarea instrumentelor, pentru retribuirea personalului auxiliar, pentru prelucrarea datelor la calculator etc., astfel încît să nu se omită nici un amănunt - mergînd pînă la cheltuieli pentru creioane şi radiere - şi finalmente să ne putem încadra în volumul fondurilor avute la dispoziţie.

Apoi, se pune problema muncii în cadrul echipei de cerce-

tare; fiecare membru trebuie să cunoască exact ce are de făcut şi ce responsabilităţi îi revin, atît din punctul de vedere al cercetării ştiinţifice, cît şi din cel administrativ-organizatoric. Dacă pentru întregirea echipei este nevoie şi de un corp de auxiliari (de exemplu, operatori de anchetă), atunci sarcinile pe această linie se măresc simţitor. In faza pregătitoare investigaţiei se va urmări cu toată atenţia problema selecţiei acestora, pe baza unor criterii sau probe stabilite, precum şi instruirea lor în legătură cu munca pe care o vor avea de desfăşurat.

Un alt set de aspecte ce se cer rezolvate cu maximum de atenţie tot în faza pregătitoare este cel legat de condiţiile de lucru şi de viaţă pe teren. De regulă, este necesar şi recomandabil să se ia legătura din vreme cu organele locale atît pentru a fi încunoştiinţate despre obiectivele cercetării (solicitînd, dacă este cazul, aprobarea desfăşurării ei), cît şi pentru a primi un sprijin preţios în privinţa cazării, a asigurării mesei etc.

Mai menţionez, în sfîrşit, necesitatea programării altor aspecte privind desfăşurarea muncii propriu-zise în teren:

- ce sarcini zilnice revin unei persoane,
- cum şi la ce intervale se efectuează controlul operaţiilor,
- locul şi datele unor eventuale şedinţe în plen ale echipei,
- cînd se efectuează predarea informaţiilor culese etc.

Iată doar cîteva din noianul de probleme organizatorice ce se ridică pentru orice cercetare sociologică de teren de mai mare amploare. Se observă, că, în mare, aceste aspecte pot fi grupate în două categorii: unele legate de activitatea de cercetare ştiinţifică propriu-zisă şi altele de asigurarea condiţiilor pentru buna ei desfăşurare. Din acest motiv, pe lîngă un "comandament" ştiinţific, care să coordoneze primele aspecte, e nevoie şi de una sau mai multe persoane cu atribuţii strict organizatorice.

1.lo. Studiul pilot

In cazul cercetărilor sociologice ce vizează probleme sau utilizează instrumente de studiu inedite, intervine, de regulă, o etapă ce se poate plasa pe o poziţie intermediară între pregătirea

cercetării şi investigaţia propriu-zisă, numită studiu pilot. Este, în fapt, un fel de "repetiţie generală", a cărei nevoie se simte mai cu seamă în cazul anchetelor. Sintetizînd la maximum funcţiile studiului pilot, putem afirma că acest moment are următoarele obiective principale:

a) testarea instrumentelor de culegere a datelor (în mod cu totul deosebit a chestionarelor);

b) familiarizarea operatorilor cu instrumentele şi, în general, cu munca de teren;

c) obţinerea de sugestii în vederea prelucrării datelor

Dacă e vorba de o anchetă, se va alege un egnation redus de persoane - de la cîteva zeci la loo - , a cărui reprezentativitate nu ne interesează decît în măsura de a putea surprinde în cadrul său situaţii cît mai diverse, eşantion care nu se suprapune, de regulă, cu cel ales pentru a fi efectiv investigat. Acestor persoane din minieşantionul de probă li se aplică chestionarul, în condiţii cît mai apropiate de cele ale cercetării. Se vor obţine astfel, cum spuneam, informaţii legate de felul în care "funcţionează" instrumentele de cercetare, se vor depista întrebările cu formulări defectuoase sau cele la care nu s-au prevăzut corect variantele de răspuns. De asemenea, vor apare informaţii asupra atitudinii subiecţilor faţă de cercetare şi vor fi puse în evidenţă acele întrebări "dificile" ce pot produce distorsionarea răspunsurilor la întreg chestionarul şi multe, foarte multe alte aspecte capabile să amelioreze forma de prezentare şi conţinutul întrebărilor.

Asupra celui de al doilea aspect, pomenit la punctul b), iarăşi nu cred că este nevoie să insist prea mult. Oricine a lucrat cu operatori de anchetă amatori ştie cîte probleme se ridică la completarea primelor chestionare, cîte greşeli apar datorită faptului că în pregătirea lor "teoretică" pentru cercetare nu se pot prevedea toate situaţiile concrete ce apar pe teren. Studiul pilot îi va ajuta să cunoască mai bine structura chestionarului, să interpreteze în mod uniform conţinutul întrebărilor , să se deprindă cu tehnica de înregistrare (codificare), să găsească formulele cele mai potrivite de deschidere a discuţiei cu subiecţii ş.a.m.d.



In sfîrşit, datele obţinute pe acest mic eşantion pot furniza sociologului o indicaţie prealabilă asupra tendinţelor ce se vor manifesta în populaţia de investigat, fapt care-l va ajuta să-şi planifice mai eficient programul de prelucrare finală a informaţiei, sugerîndu-i ipoteze suplimentare de lucru.

Se înţelege că în urma studiului pilot va avea loc definitivarea chestionarului şi o reinstruire a operatorilor, axată pe problemele ce au apărut- In plus, acest studiu va permite să se evalueze mai exact timpul necesar realizării anchetei şi sarcinile ce le poate îndeplini un operator pe parcursul unei zile de lucru. Astfel, sînt posibile corecţii în programul de cercetare pe teren.

Studiul pilot, în aparenţă, măreşte costul unei anchete prin faptul că introduce operaţii suplimentare. In fapt însă, aceste pierderi sînt înzecit compensate de cîştigurile şi avantajele care-i urmează. Intr-adevăr, odată operatorii dispersaţi în teren, orice schimbare în tactica şi strategia iniţiale devine practic imposibilă. Un chestionar tipărit şi pus în lucru nu poate fi x schimbat pe parcurs decît cu mare greutate şi cu multe riscuri Chiar dacă există posibilitatea reunirii operatorilor după prima zi de anchetă, pentru analiza problemelor practice ivite, nu există garanţa că eventualele corecturi în chestionar sînt efectuate şi respectate în continuare de toţi operatorii. Dacă dispersarea în teren este ridicată, astfel de reuniuni nu sînt posibile şi toate întrebările defectuoase vor trebuie pînă în urmă sacrificate; acelaşi lucru se va întîmpla şi atunci cînd se lucrează prin autocompletarea chestionarelor, distribuite simultan întregului eşantion.

Pe de altă parte, dacă anumiţi operatori de anchetă - cînd se foloseşte această tehnică - nu înţeleg exact sensul unor întrebări, ei vor introduce erorir sistematice a căror depistare este, evident, preferabilă înaintea cercetării propriu-zise, cînd pot fi eliminaţi şi operatorii ce se dovedesc incapabili să desfăşoare o muncă de calitate. Chiar dacă avem posibilitatea să verificăm pe teren activitatea operatorilor, pînă cînd sînt supuşi toţi unui control serios, o parte a cercetării este deja epuizată iar erorile introduse şi eventualele stereotipuri dobîndite cu greu se mai pot elimina în continuare.

Iată suficiente motive, care, deşi prezentate pe scurt şi fără pretenţia de exhaustivitate, pledează pentru respectarea acestei sugestii metodologice (prezentă mereu în manuale, dar rareori pusă în practică) conform căreia majoritatea cercetărilor sociologice au nevoie de o probă prealabilă, numită studiu pilot.

2. Valorificarea informaţiei

Dacă despre cea de a doua etapă mare a cercetării nu am găsit de cuvinţă să spun nimic în acest capitol introductiv şi general, lucrul se datorează desigur aspectelor cu totul specifice fiecărei metode şi tehnici de cercetare, astfel încît însăşi tratarea acestora se va suprapune în bună măsură cu prezentarea activităţilor de culegere a informaţiei. Voi trece deci la ultima etapă, care nici ea nu se pretează la deosebit de multe analize de ordin metodic cu caracter de generalitate.

După terminarea fazei de culegere a datelor, cercetătorul sau echipa de cercetare trebuie să procedeze la o verificare a acestora, din punctul de vedere al completitudinii, exactităţii, al modului în care sînt trecute în formularele de înregistrare, al acurateţei scrisului etc. E nevoie de acest lucru atît pentru a depista unele lipsuri şi erori strecurate în procesul de investigaţie, cît şi pentru a preveni eventualele erori ulterioare datorate unor neuniformităţi de scriere, ilizibilităţi ş.a.m.d.

Următorul pas îl constituie prelucrarea sumară a datelor. Aşa cum am spus în capitolul anterior, în marea lor majoritate, datele sociologice sînt de tip statistic şi atunci această prelucrare sumară înseamnă construcţia unor tabele simple de frecvenţe (tabelarea) însoţită de procente, medii, indicatori de dispersie etc. Aceste tabele ne vor arăta, într-o primă formă, felul în care se distribuie indivizii în clasele caracteristicilor construite, tendinţa lor de grupare sau împrăştiere în jurul valorilor mijlocii.

E de la sine înţeles că de multe ori este nevoie de o fază prealabilă, pregătitoare acestei prelucrări: postcodificări, analize de conţinut a unor texte sau întrebări deschise din chestionar etc. Dacă prelucrarea datelor se realizează pe calculator, trebuie ca informaţia să fie adusă la o formă adecvată pentru a putea fi transferată în memoria acestuia; uneori este nevoie de

transcrierea ei pe imprimate speciale,alteori poate fi transpusă pe cartelele perforate direct de pe instrument (chestionar, formular de înregistrare etc.)

c) Urmează apoi un alt pas important, şi anume prelucrarea şi analiza datelor în profunzime, adică îmbinarea variabilelor (corelaţii, asocieri, regresie) şi construirea de modele de depen-denţă. Rostul acestei analize îl constituie testarea ipotezelor explicative avansate la începutul cercetării şi obţinerea unor in-formaţii de natură predictivă asupra fenomenelor sociale. Opera-ţiile de fineţe se realizează , chiar şi pe calculator, după analiza sumară, întrucît vor trebui luate în considerare aici doar varia-bilele ce s-au dovedit relevante, capabile să dea o descriere adec-vată fenomenelor reale reflectate.

d) Pe parcursul tuturor fazelor de prelucrare a datelor şi apoi după terminarea lor, se vor degaja concluziile investigaţiei, fie sub forma unor elemente pur descriptive, fie sub formă expli-cativă sau predictivă. Toate acestea vor fi reţinute, clasificate, analizate, puse în legătură cu schema explicativă prealabilă, cu ipotezele care au ghidat cercetarea.

Ultimul act al oricărei investigaţii îl constituie redac-tarea raportului de cercetare. Folosesc aici acest termen într-un sens mai general decît cel obişnuit. De regulă, prin raport de cer-cetare se înţelege materialul întocmit de sociolog pentru benefi-ciarul care a finanţat cercetarea. Prin extensie, expresia poate fi utilizată şi dacă rezultatele se fructifică sub formă de articol, comunicare, studiu de culegere sau chiar carte. Desigur că fiecare dintre formele respective implică exigenţe specifice în redactarea şi prezentarea materialului. Intr-o carte, de exemplu, se vor accen-tua incomparabil mai mult aspectele teoretice ale problemei, se vor face incursiuni mai vaste în istoricul cercetării sale şi, evident, se vor da mult mai multe detalii decît într-un articol sau comuni-care. Dacă materialul este destinat unui beneficiar, el va trebui să aibă o prezentare grafică deosebit de sugestivă , concluziile să fie clare şi exacte, şi în general, textul să fie limpede şi concis. Tot aici se vor accentue cu precădere laturile aplicative ale studiului şi soluţiile desprinse din cercetare, în scopul ame-liorării aspectelor sociale urmărite.

Dar cred că este mai bine să schițez, după Th. Caplow (1970), criteriile ce pot fi folosite în evaluarea unui raport de cercetare. Ele vizează patru puncte esențiale:

a) Enunțul problemei studiate. Problema de studiu trebuie enunțată cu claritate în raport, astfel încît să nu existe ambiguități iar, dacă e posibil, să se prezinte ipotezele formalizate și metodele de verificare pentru acestea.

Un studiu sociologic se cuvine să poată fi evaluat și prin semnificația problemei cercetate, mergînd de la o valoare maximă, cînd soluțiile interesează întreaga masă de sociologi sau un public foarte larg, pînă la una minimă, cînd soluția este absurdă, imposibilă sau fără valoare.

In al treilea rînd, un studiu se poate caracteriza prin gradul de cuprindere a documentației existente în problema respectivă. Aprecierea studiului este aici, cum spuneam, în funcție de destinația sa, deci nu se poate pretinde întotdeauna documentația să apară cît mai complet. (E vorba de textul redactat și nu de munca de pregătire a cercetătorului!)

b) Descrierea tehnicilor de cercetare. In esență, e vorba la acest punct ca din textul raportului să reiasă faptul că:

(i) metodele, tehnicile și instrumentele folosite sînt adaptate problemei;

(ii) eșantionul, în cazul cînd se folosește, are un grad de reprezentativitate satisfăcător, cu precizarea limitelor de confidență pentru valorile calculate și a nivelului de probabilitate cu care diferiți coeficienți sînt semnificativi;

(iii) metodica cercetării poate fi reluată de către alți cercetători în scopul verificării concluziilor obținute, de unde rezultă că este nevoie de o prezentare suficient de detaliată a acesteia.

c) Prezentarea concluziilor. Indiferent de natura raportului de cercetare, concluziile trebuie prezentate în mod clar și precis, doar gradul de detaliere a lor fiind în funcție de destinația materialului scris. O prezentare a concluziilor poate fi considerată clară dacă un cititor cu formație medie este capabil să înțeleagă rezultatele sale. Nu de puține ori sociologii folosesc un limbaj prețios, care nu are finalmente alt rol decît să ascundă

slăbiciunile de conţinut.

d) Interpretarea rezultatelor. Desigur, e vorba aici de faptul că în întreg discursul argumentaţia trebuie să fie corectă şi să nu se ascundă, prin tot felul de artificii, punctele slabe ale ei. Se poate da şi o evaluare a importanţei şi utilităţii rezultatelor obţinute , din punct de vedere ştiinţific sau din cel practic aplicativ.

In sfîrşit, termin acest capitol tot cu o sugestie a lui Caplow. Marea masă de documente ce rămîne în urma unui studiu sociologic nu trebuie nici ea neglijată. Aceste documente (chestionare, listinguri de calculator, tabele manuale etc.) se cer clasificate şi păstrate în arhivă pentru situaţii în care alţi cercetători ar dori să refacă concluziile investigaţiei, cu aceleaşi proceduri de analiză sau cu altele.

Cap. III. CALITATIV SI CANTITATIV IN STUDIUL FENOMENELOR SOCIALE

Discuţiile asupra raportului calitativ-cantitativ în abordarea realităţilor sociale ocupă numeroase pagini, atît în tratatele generale de sociologie, în lucrările de filosofie socială, cît şi în studiile cu caracter strict metodologic sau metodico-tehnic. Cu siguranţă că problema este importantă din toate punctele de vedere, începînd cu cel general filosofic şi terminînd cu cel practic, legat de tehnicile investigaţiilor empirice. Pe de altă parte, este uşor, de închipuit că amplitudinea unei asemenea dezbateri ca cea stîrnită de relaţia cantitativ-calitativ este rezultatul direct al unei diversităţi enorme de poziţii, mergînd de la cele care susţin imposibilitatea cuantificării fenonenelor sociale şi pînă la cele care văd în cuantificare şi măsurare singurele operaţii capabile de a scoate aceste ştiinţe, în speţă sociologice, din starea de subdezvoltare în care, pe drept sau pe nedrept, se consideră adesea că ele se află.

In rîndurile care urmează, n-o să angajez o dezbatere cu caracter general asupra problemei enunţate în titlu, care n-ar face decît să mai adauge cîteva opinii - mai mult sau mai puţin interesante - la multitudinea celor care există deja. Voi încerca să precizez unele concepte de bază, privite în special prin prisma utilizării lor în cercetările concrete, ajutînd, sper, prin aceasta la o înţelegere mai clară a problematicii, fapt care va putea eventual constitui punct de plecare pentru unele generalizări sau discuţii epistemologico-metodologice.

Nu îmi ascund totuşi poziţia de susţinător al introducerii procedurilor cantitative în sociologie, considerînd că, în momentul de faţă, este nevoie de a sprijini din toate puterile orientările angajate pe calea aplicării în ştiinţele sociale a metodelor validate de ştiinţele pozitive. Asistăm acum, probabil ca o reacţie la excesele de empirism şi de aplicare a matematicilor fără discernămînt, la o recrudescenţă a concepţiilor privind singularitatea metodologică a ştiinţelor sociale, la o repliere a unor cercetători în sfera metodelor subiective, la atacuri virulente la adresa

încercărilor de cuantificare şi măsurare. Mă voi strădui să arăt că aspectele cantitative au un cîmp mult mai larg de manifestare în viaţa socială decît se crede în general, că noţiunea de cantitativ nu este identică cu, ci mai largă decît cea de măsurare, că prin transferul de la anumite niveluri la altele ceea ce nu este măsurabil devine măsurabil, deci că o serie de critici sînt, în esenţa lor, viciate de o înţelegere mult prea îngustă a obiectului asupra căruia se exercită.

1. Precizarea noţiunilor

Fără a intra în probleme epistemologice de mare subtilitate va fi suficient, pentru cele ce ne interesează aici, să observăm că orice ştiinţă empirică studiază un domeniu al realităţii, deci un univers de elemente a căror existenţă este presupusă a fi independentă de acţiunea subiectului cercetător, realizînd în realitatea empirică un dublu decupaj: cel referitor la obiectele materiale (în sensul larg al termenului) şi cel care priveşte o anumită gamă de caracteristici ale acestora. Obiectele materiale le voi numi obiecte-suport, spre a le distinge de cele ştiinţifice, construite de cercetător, şi le voi considera ca purtătoare a unor însuşiri, înglobînd în această noţiune şi cele de trăsătură, proprietate, caracteristică etc. Este clar că ştiinţe diferite pot porni la studierea aceloraşi obiecte, dar privite din perspectiva unor însuşiri deosebite. Aşa de exemplu, şi fizica şi chimia se ocupă de studiul atomilor, dar fiecare dintr-o altă perspectivă. In sociologie, asemenea obiecte pot fi indivizii umani, grupuri - mai mari sau mai mici - instituţii, unităţi teritoriale sau de habitat etc.

Din cele de mai sus rezultă imediat că nu este niciodată suficient, pentru a desemna obiectul unei investigaţii, să se indice obiectul material vizat, ci trebuie adăugate şi însuşirile cercetate. Astfel, nu e suficient să spunem că studiem, de pildă, un microgrup, ci va trebuie să precizăm şi perspectiva din care-l privim; ne pot interesa relaţiile interpersonale în cadrul său, frecvenţa cu care este exprimată o anumită opinie, atitudine etc. de membrii grupului , reţeaua prin care circulă informaţiile de o anu-

mită natură între indivizi şi multe alte probleme. Tot aşa, nu ne va interesa satul sau un sat în sine, ci ca o comunitate caracterizată de fenomene specifice ce se petrec în interiorul său ori în raporturile cu alte unităţi. Nici măcar în studiile cu caracter monografic nu se studiază unitatea monografiată în sine, ci prin prisma unui complex de manifestări ale ei, set care, de astă dată, se urmăreşte a fi cît mai complet posibil. Aşadar, în ultimă instanţă, corespondentul real al obiectului ştiinţific îl constituie aceste trăsături şi relaţii, caracterizînd obiectele materiale suport.

O primă - şi cea mai frecventă - accepţiune a termenilor de calitativ şi cantitativ rezultă din modalitatea de abordare, de investigare şi interpretare a însuşirirlor şi relaţiilor respective. Intr-adevăr, aceste trăsături se manifest sub forma unor stări ce pot fi percepute fie direct cu organele de simţ, fie prin intermeddiul unor aparate de genul celor din fizică, chimie, etc., fie doar mental, mai ales în cazul fenomenelor sociale, prin încadrarea lor în nişte grile socialmente constituite. Din punctul de vedere al culorii, un obiect poate fi perceput ca verde, albastru etc., din alt punct de vedere el poate fi categorisit ca mare sau mic, din altul poate fi apreciat drept urît sau frumos etc. In studiile sociologice, găsim nenumărate exemple prin care să punem în evidenţă stările unei însuşiri. De pildă, o persoană (obiectul), avînd statutul de student (însuşirea) poate învăţa la facultatea de filologie, istorie, matematică etc. (stările); un om al muncii (însuşirea) poate lucra în domeniul agriculturii, industriei, comperţului etc. (stările) ş.a.m.d.

Exemplele de mai sus sînt ilustrări tipice de însuşiri evaluate din punct de vedere calitativ, fiecare stare a însuşirii fiind deci o calitate a acesteia. Abordarea sau evaluarea cantitativă este posibilă, în principiu, atunci cînd stările trăsăturii posedă un caracter de gradualitate, cînd trecerea de la una la alta înseamnă o creştere sau descreştere a intensităţii manifestării. Această variaţie poate fi continuă sau discretă. In primul caz, trecerile de la o stare la alta sînt line, diferenţele dintre poziţii diferite dar apropiate apar greu de sesizat, deci delimitarea exactă a unor stări se face cu dificultate, practic orice

"tăietură" pe scala de variație presupunînd un oarecare grad de convenționalitate; din contră, în cazul secund, trecerile sînt brugte, diferențele dintre stări foarte clare și nete. Multe trăsături ale indivizilor umani, mai ales de natură psiho-socială, deși nu avem mijloace clare de a evidenția acest lucru, sînt considerate a avea o manifestare sub formă continuă. Astfel, gradul de satisfacție al omului în raport cu ceva este deseori apreciat a avea un interval de variație continuă, de la o stare de totală insatisfacție la una de maximum. Sau, în sociologia politică, nu o dată convingerile politice ale oamenilor sînt plasate (imaginar vorbind) pe o scală ce ar merge de la extrema dreaptă la cea stîngă. Chiar dacă practic, în astfel de cazuri, sînt evidențiate doar unele stări discrete - cîteva grade de satisfacție: mare, mică, moderată etc. - de exemplu - ele nu sînt considerate decît niște substitute, aproximări foarte imperfecte ale unei scale obiective, capabilă să evidențieze o infinitate de nuanțe.

Evaluarea cantitativă a unor asemenea însușiri poartă denumirea de măsurare, al cărei rezultat - măsura - este o expresie numerică a intensității însușirii studiate. Nu o dată măsurarea este definită drept operația prin care se atașează diferitelor stări ale fenomenului numere reale. Definiția aceasta curentă mi se pare totuși prea largă, deoarece simpla atașare de numere nu înseamnă totdeauna o măsurare, în sensul consacrat pe care-l are acest termen. De exemplu, dacă întocmim o clasificare a studenților din anul II filosofie după cursul de limbă străină pe care-l urmează, obținem 6 categorii ce pot fi numerotate astfel:

1. Engleză
2. Franceză
3. Germană
4. Italiană
5. Rusă
6. Spaniolă

Ulterior, lucrînd cu grupele respective, ele pot fi indicate doar prin simbolurile numerice acrodate, găsindu-ne deci într-o situație care se încadrează exact în definiția amintită. A spune însă că numerele atașate reprezintă o măsură înseamnă a goli termenul de măsură de orice conținut.

Prin urmare, măsurarea trebuie să constituie o operaţie mult mai complexă decît simpla atribuire de numere. In esenţă valorile acordate trebuie să se coreleze, prin mărimea lor, cu nivelul sau intensitatea însuşirii evaluate. Desigur că o regulă formală, capabilă să surprindă exact semnificaţia corelaţiei în cauză, este greu de dat, ceea ce nu înseamnă că, în practică, nu vom putea distinge, de cele mai multe ori, simplele ataşări arbitrare de numere de cele în care valorile evaluează cantitativ însuşirea. Dacă totuşi dorim o definiţie, vom putea, provizoriu şi în mod suficient pentru necesităţile cursului de faţă, să pretindem că, pentru a putea vorbi de măsurare, în mulţimea valorilor atribuite să fie posibilă realizarea cel puţin a unei operaţii aritmetice simple-scăderea - operaţia şi rezultatul ei avînd o semnificaţie şi un sens. Mai clar spus, diferenţele aritmetice dintre valorile numerice trebuie să reflecte diferenţele reale între gradele de manifestare a fenomeneului.

Un astfel de lucru nu este desigur posibil în situaţii ca cea imaginată mai sus. Numerele atribuite n-au nici măcar o semnificaţie ordinală (căci scrierea putea urma foarte bine şi altă ordine decît cea alfabetică) şi cu atît mai puţin una care să ne permită să le utilizăm în vreo operaţie matematică.

Realizarea în practică a operaţiei de măsurare presupune rezolvarea următoarelor probleme;

a) Delimitarea printr-o anumită procedură a universului obiectelor suport (spre a şti pe cine măsurăm)

b) Alegerea trăsăturii, însuşirii etc. ce trebuie supusă operaţiei de măsurare (ce măsurăm)

c) Alegerea unei unităţi de măsură sau etalon (cu ce măsurăm)

d) Definirea procedurilor concrete de aplicare a etalonului sau, altfel spus, stabilirea regulilor de atribuire a numerelor pentru obţinerea măsurii (cum măsurăm).

Cele mai mari dificultăţi de măsurare în sociologie sînt legate de ultimele două puncte mai sus enunţate. In ştiinţele naturii şi în viaţa cotidiană, pentru efectuarea celor mai multe măsurători există unităţi naturale, fizice de măsurat, iar realizarea

operației propriu-zise este tot una fizică, fie directă prin manipularea etalonului, fie indirectă prin intermediul aparatelor. Determinarea masei, a lungimii, a intensității sau tensiunii curentului electric etc. sînt cîteva exemple simple de măsurători cu unități direct manipulabile (metrul, kilogramul) sau indirect (amperul, voltul). Măsurarea intensității unei atitudini sau a gradului de adeziune la o normă socială etc. sînt cazuri tipice din științele umane în care nu avem nici unități materiale și nici instrumente bine standardizate pe care să putem citi valorile ca la voltmetru sau ampermetru. Atît etalonul cît și instrumentul (teste, chestionare) trebuie elaborate pe baza unui suport teoretic - ce nu este nici el prea solid - și pe cea a unei experiențe practice în studiul fenomenelor sociale.

De cele mai multe ori în sociologie - și nu numai aici - pentru a ajunge să măsurăm o însușire, trebuie să trecem printr-o etapă intermediară ce poartă numele de cuantificare. Măsurarea și cuantificarea nu sînt deci două noțiuni identice, cum adesea se mai prezintă; prima desemenează operația concretă prin care se obține măsura, pe cînd a doua ansamblul de operații de natură conceptuală prin care se face posibilă măsurarea. Concret, în cazul sociologiei, cuantificarea presupune o analiză aprofundată a fenomenului de măsurat, descompunerea conceptului în dimensiuni, determinarea indicatorilor, construcția indicilor și, în fine, agregarea acestora într-un indice sintetic ale cărui valori reprezintă măsura. Deci cuantificarea este o operație efectuată la nivelul teoretic și metodologic, implicînd, în esență, punerea în evidență, dezvăluirea, descoperirea - în multitudinea de determinații pe care le au fenomenele sociale - a aspectului cantitativ și indicarea formelor concrete sub care el se manifestă și poate fi perceput. Deci tot cuantificarea este cea care duce la alegerea unității de măsură și la elaborarea scalei pe care se găsesc valorile - rezultat al măsurătorii. In fond, orice măsurare e precedată de cuantificare, numai că în cazul trăsăturilor foarte simple acest lucru nu este vizibil, întrucît efortul teoretic și metodologic pentru prepararea măsurătorii este minim.

In concluzie, voi afirma că aspectul cantitativ în evaluarea manifestărilor unei însușiri, trăsături etc.apare atunci cînd

este posibilă operaţia de măsurare, în sensul mai sus prezentat. In fapt, aceasta este şi cea mai frecventată accepţiune dată cantitativului şi nu de puţine ori este şi singura despre care se vorbeşte în literatura de specialitate. Din păcate, confuzia între cantitativ şi măsurare nu este deloc benignă, căci o serie de atacuri la adresa cantitativismului se hrănesc din argumente valabile doar într-o eventuală confruntare cu abuzul de măsurare.

Intr-adevăr, abordarea cantitativă şi măsurarea nu se suprapun pentru simplul motiv că prima face apel şi la alte operaţii şi tehnici decît cele metrice. Este vorba, în speţă, de operaţia de numărare. Nimeni nu gîndeşte probabil să excludă numărarea dintre componentele cantitativului, dar, curios, foarte puţini sînt cei care fac distincţia elementară dintre numărare şi măsurare. Chiar şi un autor care se ocupă într-o carte întreagă despre problema măsurării (l-am numit pe I.Mărginean, 1982) confundă cele două chestiuni, lăsînt să se înţeleagă faptul că măsurarea ar include şi numărarea. Dar noi putem înţelege acum că în timp ce prima operaţie face apel, aşa cum s-a văzut, la manipularea însuşirirlor obiectelor, a doua (numărarea) operează cu obiectele înseie. Nu măsurăm oameni, ci înălţimea sau greutatea lor, în schimb efectiv numărăm oameni sau alte entităţi materiale purtătoare de însuşiri.

Aşadar, putem vorbi de abordare cantitativă la două niveluri: al însuşirii (măsurarea) şi al obiectului (numărarea). Inversînt perspectiva, atunci şi calitativul va apare, prin simetrie, în dublă ipostază: un gen de abordare la nivelul însuşirii şi altul la cel al obiectelor-suport. Dacă în primul caz termenul se menţine în accepţiunea sa obişnuită, cunoscută şi larg folosită, în al doilea, el pare mai greu de utilizat şi probabil ar trebui găsit unul mai adecvat. Dar cum importantă nu este inventarea de cuvinte, voi menţine tot expresia de calitativ şi aici, precizînd doar sensul în care o folosesc.

Va trebui deci să vedem care este rostul numărării, dacă şi cînd intervine ea în studiile ştiinţifice şi care este specificul acelor ştiinţe care fac apel la numărare. Intr-adevăr, problemele acestea nu sînt lipsite de sens deoarece ştim bine că ramuri întregi ale unor ştiinţe nu utilizează operaţia de numărare, decît poate în mod accidental şi pentru chestiuni marginale. Astfel pen-

tru a studia structura atomului de hidrogen, de exemplu, este suficient să se studieze un atom iar rezultatele obţinute sînt imediat transferate asupra întregii mulţimi de atomi de hidrogen din univers: la fel pentru molecule, sau substanţe complexe. Proprietăţile simple ale unor elemente, din universul fizic mai ales, ne dau această posibilitate de a face abstracţie de obiectul material purtător de însuşiri şi de a generaliza o anumită proprietate, relaţie sau structură la toate obiectele de aceeaşi specie.

Uneori însă astfel de generalizări nu sînt posibile; însuşirile prezintă variaţii mari de la individ la individ, raporturile şi conexiunile depind de purtătorii materiali ai elementelor ce se combină. Atunci va fi nevoie să numărăm; să vedem cîţi indivizi au sau, mai exact , ce proporţie din rîndul lor are un comportament sau altul, ce proporţie se află într-o stare sau alta etc. Legităţile detectate în astfel de situaţii vor fi de tip probabilist iar în abordarea efectivă a realităţii se va face apel la instrumentul de analiză numit statistică. Nu este greu să observăm că aceasta este situaţia pentru majoritatea fenomenelor sociale; nu este greu deci să înţelegem de ce ştiinţele sociale fac apel în mod curent la statistică; nu este greu să intuim că în studiul fenomenelor sociale aspectul cantitativ legat de numărare nu este unul secundar, neglijabil, ci el are în realitate tot atîta importanţă ca şi cel legat de măsurare.

Din cele spuse pînă aici, rezultă deci, că pe de o parte, însuşirile obiectelor pot fi descrise fie de o manieră calitativă, fie de una cantitativă. Pe de altă parte, o însuşire poate fi urmărită, în formele sale de manifestare, fie prin inventarierea obiectelor suport - adică încercînd să se determine ce fracţiune din "universul" de obiecte studiat posedă cutare sau cutare formă de manifestare a caracteristicii - fie, prin analiza unui singur obiect, care este capabil să reprezinte corect o întreagă clasă de obiecte identice. Vom avea deci patru tipuri fundamentale de metode de analiză, considerînd simultan însuşirea (primul termen) şi obiectul (al doilea):

a) Metode calitativ-calitative
b) Metode cantitativ-calitative
c) Metode calitativ-cantitative
d) Metode cantitativ-cantitative

Primul gen cuprinde metodele care abordează trăsătura din punct de vedere calitativ şi nu ia în considerare întinderea sau frecvenţa cu care se realizează diferitele stări ale trăsăturii respective. Al doilea tip încearcă să măsoare intensitatea manifestărilor, iarăşi făcînd abstracţie de obiectul concret. Al treilea este similar primului şi al patrulea celui de al doilea, introducînd, în plus, fiecare şi obiectul ca entitate ce contează în analiză.
Aceste două din urmă metode sînt deci cele care fac apel la statistică, iar distincţia dintre ele revine la distincţia ce se face în statistică între procedurile de analiză calitativă şi cantitativă sau nonparametrice şi parametrice.

Abordarea cantitativă din unghiul obiectului este impusă de faptul că fenomenul studiat nu posedă, aşa cu, spuneam, un grad de invarianţă suficient de ridicat, în raport cu diferite condiţii, astfel încît el să se menţină constant de la un individ la altul. Dacă ştim , de exemplu, că un om este "muncitor", obţinem o informaţie (suplimentară faţă de cea conţinută în definiţie) mult mai slabă decît dacă spunem că un atom este de oxigen. Despre atomii de oxigen se poate afirma că toţi au masa atomică 16, despre muncitori că au în medie o retribuţie de 25oo lei (de pildă). Pentru formarea moleculei de apă, un atom de oxigen intră întotdeauna în combinaţie cu doi atomi de hidrogen; muncitorul, pentru întemeierea unei familii, îşi alege, cel mai adesea, un partener dintr-un mediu social apropiat etc. Iată tipuri de aserţiuni care se folosesc într-un caz sau celălalt.

Abordarea nestatistică se întîlneşte deci în foarte multe domenii ale ştiinţelor naturii, deşi chiar şi cele mai dezvoltate dintre acestea fac apel la instrumentul statistic. Ea se foloseşte curent şi în ştiinţele socio-umane, mai cu seamă atunci cînd, neavînd de a face cu o identitate a obiectelor, se caută pentru analiză unul tipic sau reprezentativ pentru o anumită clasă de obiecte. Aşa, de pildă, studiind capitalismul în Anglia, Marx a abordat un caz tipic sau clasic sau un model de sistem social capitalist.

Intre această din urmă formă de analiză şi cea statistică nu este o linie de demarcaţie riguros trasată . Alegerea cazului (cazurilor) tipic (tipice) se poate baza pe o analiză prealabilă a

masei de unităţi, tot aşa cum un studiu statistic poate urma unuia de caz. Această îmbinare de metode se practică în mod curent şi cu succes în cadrul ştiinţelor socio-umane, putînd duce la rezultate foarte bune. Analiza de caz permite studierea fenomenelor în adîncimea lor, descrierea şi explicarea conexiunilor între diferite însuşiri ale obiectului, urmărirea în timp a manifestărilor acestora etc. Analiza statistică a unei populaţii desprinde ceea ce este general, repetabil, măsoară frecvenţele cu care apar anumite stări, determină tendineţele de evoluţie, stabileşte regularităţile care se observă numai în cazul unei mulţimi de indivizi (fenomenele de masă). Aşadar, nu numai că cele două tipuri de metode nu se exclud, nu se opun unul altuia, cum se mai susţine uneori, dar ele se dovedesc a fi complementare, susţinîndu-se reciproc în investigaţiile empirice.

2. Scale de evaluare

Reluînd acum problematica din unghiul de vedere al trăsăturii, voi încerca să schiţez sumar care sînt instrumentele cele mai des utilizate pentru evidenţierea formelor de manifestare ale acesteia, stărilor în care se poate găsi sau nivelului de intensitate manifestat. Ajungem astfel la ceea ce în litetura metodologică poartă denumirea de scale de măsură, termen care mi se pare inadecvent, pentru motive care vor fi înţelese imediat, motive legate de utilizarea abuzivă a cuvîntului "măsură"; prefer, de aceea, denumirea de scale de evaluare. Tipurile fundamentale de scală sînt în număr de patru - în cadrul lor putîndu-se găsi şi unele subtipuri specifice-, mergînd de la o evaluare pur calitativă spre una pur cantitativă, în ordinea în care vor fi prezentate mai jos.

Scala nominală. Este instrumentul folosit în realizarea oricărei clasificări. Scala nominală evidenţiază direct stările însuşirii urmărite, stări care formează treptele sau elementele scalei. Cea mai simplă scală nominală este cea cu două trepte, scala dihotomică. De exemplu, pentru însuşirea "sex", există două stări. "masculin" şi "feminin", care constituie elementele scalei respective. Pentru însuşirea "ocupaţie", putem construi o scală cu următoarele elemente: "agricultor", "muncitor", "tehnician", "intelectual", "funcţionar", "alte ocupaţii".

Este evident că elementele scalei nominale, spre a fi identificate, poartă anumite etichete, sînt marcate de unele simboluri. Acestea, ca în exemplele de mai sus, pot fi cuvinte prin care să se sugereze conţinutul noţiunii prin care se constituie o treaptă a scalei sau pur şi simplu nişte simboluri abstracte, fără legătură cu acest conţinut, cum ar fi litere sau numere. Subliniez că eventuala folosire de numere pentru simbolizare nu ne dă dreptul să facem apel la nici o proprietate matematică a lor.

Aşadar, elementele scalei nominale sînt stări pur calitative, distincte, în care un obiect sau altul se poate afla. Singurele cerinţe formale ridicate de scala nominală sînt cele ale oricărei clasificări: orice individ să poată fi plasat pe o treaptă şi numai pe una a scalei. Dacă abordarea la nivelul obiectului suport este cantitativă (statistică), atunci elementele scalei nominale devin clase în care se grupează indivizii, în funcţie de însuşirea studiată. Problema practică cea mai importantă este elaborarea unui criteriu de decizie cu ajutorul căruia fiecare individ să poată fi plasat în una şi numai în una dintre clase.

Scala ordinală. Este un gen mai complex de scală decît cea precedentă, în sensul că posedă toate însuşirile scalei nominale, avînd , în plus, proprietatea că între elementele ei se stabileşte o relaţie de ordine. Spre deosebire deci de scala nominală, unde stările însuşirii erau evidenţiate prin elemente ale căror poziţii în scriere erau indiferente, la cea ordinală treptele scării respectă o anumită ordine, reflectînd o ierarhie între stări, dată, îndeosebi, de intensitatea cu care însuşirea evaluată se manifestă. Prin urmare, scala ordinală suprinde de o manieră calitativă o manifestare graduală a fenomenelor. Se pleacă cel mai adesea de la supoziţia că fenomenul în sine are o variaţie cantitativă şi că doar formele subtile de manifestare a lui, limitele cunoaşterii noastre, imprecizia instrumentelor de înregistrare ne fac să nu-l pute surprinde în această manieră.

Dacă notăm elementele (treptele) scalei cu litere mari: A,B,C...., atunci relaţia dintre ele poate fi scrisă:

şi se poate citi: "A mai mic decît B", "A se află înaintea lui B", "poziţia A este inferioară poziţiei B", "starea A este mai puţin intensă decît B" ş.a.m.d.

Ca orice relaţie de ordine, şi aceasta trebuie să îndeplinească unele condiţii fundamentale:

a) Orice ar fi două elemente ale scalei, A şi B, vom avea $A < B$ sau $B < A$.

b) Intre două elemente <u>diferite</u> ale scalei A şi B, nu putem avea simultan relaţiile $A < B$ şi $B < A$. In cazul că s-ar întîmpla aşa ceva, elementele A şi B trebuie să coincidă.

c) Dacă $A < B$ şi $B < C$, atunci $A < C$. Este celebra proprietate de tranzitivate a relaţiei de ordine.

Exemple de scale ordinale folosite în sociologie se pot da foarte uşor. Multe întrebări în cadrul chestionarelor au variante de răspuns ce se constituie în elemente ale unor asemenea scale. De pildă, la întrebări de tipul: "Cum apreciaţi emisiunea...., prezentată în ziua ... la televiziune ?" apar, în chestionare, răspunsuri precodificate de forma:

1. Foarte bună
2. Bună
3. Mediocră
4. Slabă
5. Foarte slabă

formînd o scală cu cinci trepte, în scopul de a evidenţia opinia celui chestionat.

Adesea elementele scalei se simbolizează prin cifre. Aci, spre deosebire de cazul precedent, cifrele nu pot fi ataşate, oricum, ele trebuind să respecte relaţia de ordine dintre stări. Putem numerota treptele scalei aşa cum am făcut-o mai sus, în exemplul dat, ori putem să le numerotăm invers dar nu şi în alt mod. Cu toate că scala ordinală încearcă să reflecte ceva din aspectul cantitativ, cu toate că numerele ataşate treptelor reflectă ordinea acestora (deci facem apel la relaţia de ordine dintre numere) ele - numerele - nu pot fi interpretate în termeni de măsură deoarece nu reproduc nimic din aspectul cantitativ presupus; scala ordinală rămîne o scală de tip calitativ, chiar dacă reprezin-

tă, față de cea nominală, deja un pas spre abordarea cantitativă.

Scala de intervale. Este, în esență, rezultatul cuantificării fenomenului studiat, deci o scală de măsură. Elementele scalei se reprezintă în mod obligatoriu prin numere (reale), care, de data asta, nu mai joacă doar rolul de simboluri, ci reflectă efectiv magnitudinea manifestării însușirii. Specificul acestora scale, în raport cu următorul tip, îl reprezintă un grad ridicat de convenționalitate în construcție, în sensul că, pe lîngă alegerea unității de măsură (a etalonului) - unitate care este totdeauna convențională -, cercetătorul fixează după bunul său plac și originea scalei. Mai exact spus, asemenea scale nu au un punct zero impus obligatoriu de manifestarea fenomenului.

Pentru a înțelege mai ușor specificul scalei de intervale, majoritatea manualelor dau ca exemplu, pe bună dreptate, o scală de măsură a temperaturii, de pildă scala Celsius. Aici punctul zero e ales convențional - temperatura de îngheț a apei -, el putînd fi ales la orice alt nivel (și este efectiv ales la altul pe scara Fahrenheit).

Restricția fundamentală ce caracterizează scala de intervale este că două valori măsurate cu această scală nu se pot raporta una la cealaltă. Spre exemplu, dacă astăzi la ora 15 la București sînt 3o^{o}C iar la Brașov 15^{o}C , nu este permis să se spună că la București aerul are o temperatură de două ori mai mare (sau că este de două ori mai cald) decît la Brașov. Că lucrurile stau așa o simțim fiecare dintre noi, și de obicei, nu facem astfel de greșeli de apreciere. Acum putem să și înțelegem lucrurile perfect, fiind suficient să ne imaginăm că punctul zero ar fi fost ales la nivelul de lo^{o}C. Atunci, la București am fi avut o temperatură de 2o de grade iar la Brașov una de 5 grade și, prin urmare, dacă încălcăm restricția, ar trebui să,spunem că la București e de patru ori mai cald ca la Brașov. Or, este absurd ca, în funcție de scală, același fenomen să poată fi atît de divers interpretat.

Comparații în termeni de rapoarte se pot face pe scala de intervale numai între diferențele dintre valori. Astfel, dacă în aceeași zi, la Cluj sînt 2o^{o}C, putem spune că diferența de temperatură dintre București și Cluj (3o^{o}C - 2o^{o}C = lo^{o}C) este de

două ori mai mare decît diferența de temperatură între Cluj și Brașov ($20^oC - 15^oC = 5^oC$). Oriunde am muta originea și oricum am schimba unitatea de măsură, acest raport rămîne constant. Astfel, punînd, ca mai sus, zero în locul lui 10^oC, la Cluj vom avea 10 grade. Pe noul termometru, diferența dintre București și Cluj va fi 10 grade (20- 10) iar între Cluj și Brașov 5 grade (10 - 5) și raportul diferențelor se păstrează.

Această caracteristică a scalelor de intervale poate fi matematic precizată în felul următor: orice scală de intervale poate fi transpusă în alta printr-o transformare lineară. Adică, dacă notăm cu x valorile pe care variabila le ia pe o scală, ele pot fi transformate în alte valori, y cu ajutorul formulei:

$$y = a + bx \qquad (a \neq 0, \quad b \neq 0)$$

Coeficientul a ne arată cum se schimbă originea de la trecerea lui x la y ($x = 0 \Rightarrow y = a$), iar coeficientul b cum se modifică unitatea de măsură (etalonul).

Fie acum două valori diferite ale lui x : x_1 și x_2. Transformînd scala, vom obține corespunzător două valori pentru y:

$$y_1 = a + bx_1$$

$$y_2 = a + bx_2$$

Se poate demonstra ușor că raportul x_1/x_2 este diferit de y_1/y_2, adică:

$$\frac{a + bx_1}{a + bx_2} \neq \frac{x_1}{x_2}$$

Intr-adevăr, presupunînd că cele două rapoarte ar fi egale, ar rezulta:

$$(a + bx_1).x_2 = (a + bx_2).x_1$$

sau

$$ax_2 + bx_1x_2 = ax_1 + bx_1x_2$$

deci

$$ax_2 = ax_1$$

şi fiindcă <u>a</u> este diferit de zero: $x_2 = x_1$

ceea ce este în contradicţie cu presupunea că x_1 e diferit de x_2

In schimb, raportul diferenţelor se conservă. Pentru demonstraţie, să mai luăm o valoare distinctă x_3, căreia îi corespunde:

$$y_3 = a + bx_3$$

Să calculăm diferenţele:

$$y_2 - y_1 = a + bx_2 - a - bx_1 = b(x_2 - x_1)$$

$$y_3 - y_2 = a + bx_3 - a - bx_2 = b(x_3 - x_2)$$

Deci

$$\frac{y_3 - y_2}{y_2 - y_1} = \frac{b(x_3 - x_2)}{b(x_2 - x_1)} = \frac{x_3 - x_2}{x_2 - x_1}$$

ceea ce trebuia demonstrat.

<u>Scala de rapoarte.</u> Reprezintă un instrument de măsură mai puternic decît scala de intervale deoarece permite raportarea măsurilor obţinute. Acest lucru rezultă din unicitatea punctului de origine (existenţa unui zero natural). De pildă, în măsurarea temperaturii o scală de rapoarte este scala Kelvin a temperaturilor "absolute". Singurul element convenţional al scalei de rapoarte este unitatea de măsură. Formula de transformare a unei astfel de scale în alta de aceeaşi natură este:

$$y = ax$$

Formula reflectă tocmai faptul că printr-o astfel de transformare are loc schimbarea unităţii de măsură: la o unitate a lui <u>x</u> vor corespunde <u>a</u> unităţi pentru <u>y</u>, căci dacă $x = 1$, rezultă că $y = a$. Evident aici:

$$\frac{y_1}{y_2} = \frac{ax_1}{ax_2} = \frac{x_1}{x_2}$$

Exemple de astfel de nivele de măsură se pot da cel mai simplu din activitatea cotidiană. Lungimea unui corp, indiferent că e măsurată în metri, yarzi, coţi etc. presupune o scală de rapoarte, căci zero corespunde mereu aceleaşi situaţii: lipsa însuşirii respective.

In domeniul sociologiei şi al ştiinţelor sociale, în general, atît folosirea scalelor de intervale cît şi a celor de rapoarte este mai puţin frecventă decît în cazul ştiinţelor naturii, unde însuşirile se pretează mai uşor la măsurători. Totuşi astfel de evaluări nu lipsesc nici în ştiinţele umane, ba aş putea zice că ele sînt mai numeroase decît mulţi îşi închipuie. Intr-adevăr, cînd se vorbeşte de astfel de scale, se au în vedere aproape exclusiv însuşiri ale indivizilor umani cum ar fi inteligenţa, opinii, atitudini etc. In cazul unor asemenea însuşiri este drept că puţine se pot măsura cu adevărat. Există cîteva variabile simple ca vîrsta, retribuţia, nivelul de şcolarizare ş.a. ce pot fi măsurate relativ uşor, chiar pe scară de rapoarte. In rest, mai cu seamă pentru opinii şi atitudini, întîlnim numeroase încercări de a construi scale de măsură, existînd deci o "tehnologie" specială - dezvoltată de psihologie şi psihosociologie - avînd un asemenea obiectiv. Metode sofisticate, legate de nume celebre ca Guttman, Thurstone sau Likert, au fost imaginate spre a construi scale de intervale sau de rapoarte pentru a măsura asemenea însuşiri. Nu mă voi ocupa de ele aici, căci am specificat de la început că voi accentua doar problematica specifică sociologiei, or, cum spuneam, aceste proceduri apar cu precădere în psihologia socială. Pe de altă parte, numai tratarea lor spre a fi corect înţelese şi aplicate ar lua probabil un volum egal cu jumătate din cel al cursului de faţă.

Cînd spuneam că există şi în sociologie mai multe posibilităţi de măsurare decît se întrevăd la o privire superficială, mă gîndeam la o altă chestiune şi nu la aceste scale. Ar trebui observat că pentru sociolog unităţile de analiză nu sînt întotdeauna (şi poate nici cel mai adesea) indivizii umani, ci grupurile

de indivizi, unitățile teritoriale, instituțiile etc. Or, cum voi arăta imediat, o serie de atribute ale acestora se lasă cu ușurință măsurate.

Pentru ilustrare, mă voi opri la un singur aspect care mi se pare esențial, nu numai pentru că pune în evidență varii posibilități de măsurare în sociologie, ci pentru că face o legătură directă între cele două aspecte ale cantitativului - măsurare și numărare -, aspecte pe care inițial le-am delimitat categoric, urmînd acum să arăt că este posibilă și o trecere de la unul la altul. Sintetic spus, ceea ce vreu să scot în evidență aici este constatarea că anumite proprietăți grupale (definite deci la nivelul grupului de indivizi) sînt măsurabile în virtutea faptului că ele au la bază numărarea în interiorul fiecărui grup, deci la un nivel inferior celui la care se realizează analiza.

Astfel, dacă unitatea de studiat este familia, putem defini o trăsătură a acesteia - s-o numim dimensiunea familiei - prin numărul de indivizi ce o compun. Vom obține astfel familii cu o persoană, cu 2, cu 3 etc. Aceste valori reprezintă evident o măsură a dimensiunii familiei și ele sînt elementele unei scale de rapoarte. Există un zero natural și se poate spune că o familie este de două sau de trei ori mai numeroasă decît alta.

Incă un exemplu, tot atît de simplu, poate fi următorul. Să presupunem că ne interesează grupele de studenți dintr-o universitate iar proprietatea studiată este promovabilitatea în sesiunea din vară. Cu ajutorul numărării, putem stabili, la nivelul fiecărei grupe, procentul studenților promovați în vară la toate examenele. Evident că aceste procente sînt o măsură a promovabilității în cadrul grupelor și că valorile obținute merg, în principiu, de la 0 (nici un promovat) la loo (toți promovați), realizînd finalmente din nou o scală de rapoarte.

Această pistă deschisă, adică sesizată ideea de construcție a acestor scale, este clar că numărul lor se poate înmulți fără dificultate și este tot atît de limpede că nu rareori lucrăm cu scale foarte puternice fără să ne dăm seama de acest lucru. Ideea vine deci să sprijine o alta și anume că demersurile cantitative în abordarea fenomenelor sociale nu sînt deloc excepții. Numărarea, ca un prim nivel apoi măsurarea unor însușiri individuale și, în fine, măsurarea unor însușiri de grup, avînd la bază

numărarea, sînt tot atîtea ipostaze ale cantitativului şi probabil nu singurele.

Indiferent de opiniile pe care teoreticienii le emit în legătură cu raportul cantitativ-calitativ, indiferent poate chiar de propria-i concepţie, sociologul angajat în cercetarea empirică va avea mereu tendinţa de a utiliza scale de evaluare cît mai puternice de preferinţă scale de măsură, deoarece aceste instrumente de măsură au virtuţi analitice superioare, îi deschid posibilităţi multiple de studiere a raporturilor dintre fenomene. De aceea apare uneori tendinţa - şi atunci criticile calitativiştilor sînt pe deplin justificate şi la obiect - de a "silui" fenomenele sociale, forţîndu-le să intre în cadrele strîmte ale unui instrument de măsură. Două puncte nevralgice mi se pare că trebuie relevate în acest sens:

a) aplicarea scalelor ordinale acolo unde fenomenul nu manifestă o variaţie graduală;

b) aplicarea scalelor de intervale şi rapoarte în cazul în care variaţia graduală există dar măsurarea e inconsistentă, rezultat al unei cuantificări cu un grad excesiv de convenţionalitate şi al lipsei oricărui temei de a acorda valori manifestărilor fenomeneului.

Această din urmă tendinţă poate fi uşor detectată în multe lucrări sociologice apărute în anii din urmă la noi. Fascinaţi de posibilităţile multiple de a "jongla" cu valori şi de tehnicile matematice complexe bazate pe scalele metrice, sociologii recurg la asemenea scale fără prea mult discernămînt. O întrebare de opinie cu 5 sau 7 variante de răspuns ierarhice (de genul: foarte mult, mult etc.) e automat ransformată într-o scală de măsură, dîndu-se variantelor valori de la 1 la 5 ori de la 1 la 7, valori cu care se calculează apoi medii, indici de dispersie, coeficienţi de corelaţie ş.a.m.d. Tentaţia este uşor de înţeles, dar şi în utilizarea noţiunii de măsură se cere .... măsură!

Revenind la problematica scalelor, să precizăm că acestora, ca de altfel oricărui instrument de observare a realităţii empirice , nu trebuie să le lipsească două proprietăţi fundamentale: validitatea şi fidelitatea.

O scală este validă dacă evaluează (măsoară) ceea ce

urmăreşte să măsoare. E vorba aici, în fond, de raportul dintre conţinutul conceptului corespunzător însuşirii şi mijloacele de identificare construite. Dacă măsor inteligenţa, să zicem, prin rezultatele pe care oamenii le-au avut în şcoală la matematică - ceea ce, de altfel, unora le poate părea destul de normal - atunci e aproape sigur că nu mi-am ales un instrument valid deoarece noţiunea de inteligenţă acoperă mult mai multe aspecte decît eventualele aptitudini pentru matematică. In ultimă instanţă, determinarea validităţii revine la o analiză aprofundată a conceptelor şi a modului de operaţionalizare a lor. Exemplul cu inteligenţa este foarte sugestiv pentru că redă o caracteristică mai generală a multor noţiuni din domeniul ştiinţelor sociale: pluridimensionalitatea lor, bogăţia de sensuri şi semnificaţii. S-ar putea spune că practic orice concept din domeniul ştiinţelor noastre suferă o sărăcire, o limitare cînd este operaţionalizat şi deci cînd însuşirea este supusă măsurătorilor. Conştientizîndu-ne acest lucru, nu trebuie să-l considerăm ca un obstacol insurmontabil, ci doar ca un permanent semnal care să ne atragă atenţia asupra limitelor cunoaşterii ce utilizează asemenea metode, asupra caracterului parţial al concluziilor obţinute şi deci asupra necesităţii de a desfăşura un permanent travaliu spre desăvîrşirea instrumentelor de analiză.

O scală este <u>fidelă</u> dacă asigură rezultate constante atunci cînd se aplică în mod repetat aceluiaşi fenomen. Fidelitatea nu presupune, evident, valididtatea căci pot obţine mereu cu un instrument aceleaşi rezultate fără ca el să-mi măsoare ceea ce urmăresr eu. Fidelitatea este deci o condiţie minimă, necesară, pe care trebuie s-o îndeplinească orice instrument, fără a fi însă şi una suficientă. Testarea fidelităţii reprezintă o cerinţă elementară pentru orice scală construită. Nici acest lucru nu este însă chiar atît de simplu în ştiinţele sociale, care au de a face cu fenomene schimbătoare şi deci nu există certitudinea că la o repetare a scalei se evaluează acelaşi lucru ca în momentul iniţial.

In încheierea acestor rînduri, ţin să precizeze, acum după ce, prin succinta prezentare a scalelor, s-au înţeles poate mai bine aspectele abordare în primul paragraf, că insistenţa mea asupra virtuţilor analizei cantitative a fenomenelor sociale nu urmăreşte negarea rolului pe care metodele calitative îl pot juca

în sociologie. Este limpede că nici o ştiinţă nu urmăreşte, în fond, eliminarea calitativului, ci doar a subiectivismului, care în ştiinţele umane se ascunde adeseori în spatele unor abordări calitative superficiale şi se maschează sau chiar se justifică prin pretinsa imposibilitate a cuantificării fenomenelor sociale. Extrema varietate a acestor fenomene implică şi o mare diversitate în posibilităţile de măsurare a lor, deci multiple situaţii de împletire a cantitativului cu calitativul, demersuri diferite dar care se completează reciproc.

Complexitatea unor fenomene sociale nu trebuie să ne descurajeze în încercarea de analiză riguroasă a lor şi să ne arunce în braţele unui impresionism subiectivist. Nu pot desigur măsura, de pildă, satisfacţia oamenilor într-o situaţie oarecare aşa cum măsor intensitatea curentului electric. Pot însă operaţionaliza conceptul de satisfacţie, pot construi o scală cu două, trei, patru sau mai multe trepte de satisfaceţie şi pot număra cîţi indivizi ai populaţiei studiate intră în fiecare clasă obţinută. Astfel, pot afla ce proporţie se află spre partea pozitivă şi ce proporţie spre cea negativă a scalei, văzînd care situaţie este predominantă; pot repeta cercetarea şi afla în ce direcţie evoluează fenomenul; pot compara situaţia populaţiei mele cu a alteia; pot găsi subclase ale populaţiei în care distribuţia pe scală e diferită, ceea ce înseamnă că am detectat un fenomen ce-mi poate explica parţial pe cel studiat etc. Aceştia sînt cîţiva paşi modeşti, dar fermi în acelaşi timp, pe drumul ce duce la cunoaşterea obiectivă a realităţii sociale, la posibilitatea sociologului de a afla ceea ce se ascunde în spatele unor aparenţe, de a transcede dincolo de "explicaţiile" sociologiei vulgare de care fiecare dintre noi sîntem impregnaţi.

3. Aplicaţiile matematicii şi statisticii în sociologie

Aplicarea matematicii în ştiinţele socio-umane a devenit astăzi o realitate de care nimeni nu se mai poate îndoi. Economia, demografia, lingvista, psihologia şi alte discipline, printre care şi sociologia, beneficiază, în măsură mai mare sau mai mică, în funcţie de nenumăraţi factori obiectivi sau subiectivi, de acest

puternic instrument de cunoaştere care e limbajul matematic. Adversarii de principiu ai utilizării matematicii în ştiinţele sociale - cel puţin în rîndul celor care practică efectiv aceste ştiinţe - sînt tot mai puţini; singurele critici cu adevărat obiective ce pot fi aduse respectivei utilizări sînt legate de abuzul de matematizare şi cuantificare, de încercările unora de a prezenta, într-o formă prezenţioasă şi într-un limbaj adesea inaccesibil majorităţii sociologilor şi publicului larg, nişte rezultate minore, care ar fi putut fi expuse tot aşa de bine şi sub o formă mult mai simplă, în limbajul obişnuit.

Mersul înainte al ştiinţelor sociale nu se poate realiza, după părerea mea - şi, desigur, nu-mi revendic aici nici un dram de originalitate - pe altă cale decît cea parcursă de ştiinţele naturii. Or, cum se ştie şi cum s-a subliniat adesea, acestea din urmă au progresat în măsura în care au fost capabile să facă apel la instrumentul matematic. De la Bacon încoace, numeroşi oameni de ştiinţă şi filosofi au subliniat corelaţia ce există - sau trebuie să existe - între spiritul ştiinţific şi cel matematic. Sociologia, ca nici o altă ştiinţă, în măsra în care se vrea ştiinţă, iar termenul acesta nu e golit de orice conţinut, nu poate face excepţie de la o regulă ce s-a manifestat atît de puternic pînă acum.

Analizînd situaţia de fapt, existentă azi în sociologie, putem constata uşor că aparatul matematic este aici mai puţin prezent, iar rezultatele obţinute pe această cale mai puţin spectaculoase, decît în alte ştiinţe, chiar sociale. Incercări de a explica această situaţie sînt destule, mergînd de la reproşul făcut matematicienilor că nu dezvoltă suficient acele capitole din domeniul lor care se pretează la aplicaţii în sociologie şi pînă la cele adresate sociologilor că sînt insensibili la spiritul matematic, rămînînd la o formaţie umanistă desuetă sau pur şi simplu că nu sînt capabili să priceapă nimic din matematică. In cadrul acestei discuţii revine mereu, ca un laitmotiv, problema enunţată în paragrafele precedente, şi anume cea a raportului dintre cantitativ şi calitativ. Se argumentează că fenomenele sociale se pretează greu la cuantificări şi deci acesta ar fi motivul pentru care matematicile, prin excelenţă ştiinţe ale cantităţii, nu-şi găsesc cîmp

suficient de larg de aplicare în sociologie. Se argumentează şi invers, că matematica a încetat de a folosi doar elemente cantitative, că laturile sale calitative ar fi cele preponderente azi şi deci că sociologii n-au decît să caute şi vor găsi ceea ce le este necesar.

O discuţie şi o polemică la acest nivel atît de abstract mi se pare cu totul inutile. După ce am precizat cît de cît înţelesul termenilor de cantitativ şi calitativ, pentru a sugera unele răspunsuri la întrebări pertinente, se cuvin cîteva cuvinte asupra înţelesului expresiei aplicarea matematicii în sociologie. Voi observa, mai întîi, că uneori se foloseşte ca termen echivalent cel de matematizare a sociologie, dar cu o frecvenţă totuşi mai mică, poate şi pentru că este puţin mai tare, sugerînd o infiltrare mai puternică a matematicii în sociologie, golirea acesteia de conţinutul său concret.

S-a subliniat adesea dinstincţia de esenţă ce există între ştiinţele logico-matematice şi ştiinţele nomotetice sau ale realului sau empirice, indiferent cum le vom numi. Dacă această distincţie este corectă - şi cred cu toată convingerea că lucrurile stau aşa -, atunci rezultă, printre altele, că matematica nu poate fi altceva pentru sociologie decît o metodă, un instrument, un auxiliar, un limbaj. Cu alte cuvinte, nu e nici un pericol de dispută între aceste ştiinţe pentru vreo chestiune teritorială. Asemenea diferende pot apare doar între ştiinţele realului care se învecinează pe porţiuni comune; matematica nu se învecinează cu nici una, fiind, în acelaşi timp, în relaţie cu toate.

Dacă acceptăm concepţia curentă conform căreia orice ştiinţă nomotetică este structurată pe două niveluri principale - cel empiric şi cel teoretic -, atunci putem constata că matematica îşi găseşte posibilităţi de aplicare pe ambele paliere. La nivelul cercetării sociologice empirice, matematica permite o analiză a datelor sociologice culese, o prelucrare a lor, şi mai ales, construirea unor modele şi scheme explicative.

Pentru exemplificare, să presupunem că între mai multe variabile am stabilit o relaţie funcţională de genul:

$$y = f(x_1, x_2, \ldots, x_n)$$

i studiere matematică poate contribui la realizarea citorva cognitivi:

a) Evidenţierea raporturilor cantitative existente între ..bila dependentă, y, şi cele independente.

b) Detectarea influenţei specifice a fiecărui factor x ..ra lui y . Astfel, stabilind că pe un domeniu oarecare funcţia ..e derivabilă în raport cu x_1 şi că, de exemplu,

$$\frac{\partial f}{\partial x_1} > 0$$

..unci vom putea spune că, în limitele domeniului unde se reali..ează inegalitatea de mai sus, o creştere a lui x_1 antrenează o ..reştere a lui y, ceilalţi factori rămînînd constanţi. Si în mod ..imilar se poate proceda cu toate variabilele independente.

c) Formularea unor previziuniîn legătură cu y. Cunoscînd ..endinţa de variaţie a fiecărui x, prin combinarea acestora în ..uncţia f, va rezulta tendinţa lui y.

La nivel teoretic, aplicaţiile matematicii în cadrul unei ştiinţe înseamnă, în esenţăm formalizarea limbajului său. Cît priveşte sociologia, consider că deocamdată această problemă nu se pune, cel puţin la nivelul teoriei generale. In cadrul anumitor teorii parţiale sau de rang mediu, limitate la o sferă de fenomene simple, s-au realizat încercări de formalizare, dar ele nu au reuşit încă să-şi demonstreze pe deplin viabilitatea.

Se poate, cred, aprecia că, în momentul de faţă, încercările de aplicare a matematicii în sociologie au un efect benefic pentru ambele discipline. Sociologia, chiar dacă nu realizează profituri explicative deosebite, va fi pozitiv influenţată în dezvoltarea sa cel puţin pe linia precizării conceptelor. Se cunoaşte rigoarea limbajului matematic şi, în acelaşi timp, imprecizia ce domină în cel sociologic. Intîlnirea celor două discursuri nu se poate face atîta vreme cît nu se aruncă o punte peste această adevărată prăpastie. Or, puntea trebuie să fie aruncată de pe tărîmul sociologiei căci matematica nu poate renunţa, fără a se nega pe ea însăşi, la exactitatea conceptelor sale.

Pe de altă parte, matematica poate benefia de aplicaţii în domenjul social, în scopul extinderii şi diversificării teoriilor

sale. Este adevărat că dezvoltarea de pînă acum a domeniilor matematice a avut drept stimulent extrinsec aplicaţiile în domeniul ştiinţelor naturii sau tehnice. De aceea e plauzibil că, cel puţin parţial, ea şi-a concentrat activitatea spre a răspunde cît mai bine acestor cerinţe, influenţîndu-se astfel structura, raportul de forţe între diferitele ramuri ale sale. Nu trebuie totuşi exagerat acest aspect, nu trebuie supraestimată o atare influenţă. Multe descoperiri matematice n-au avut nici o legătură practică imediată, iar, dacă şi-au găsit-o, asta s-a întîmplat peste decenii sau secole. Domeniul matematic este relativ autonom, avînd o logică a sa internă de dezvoltare.

Legat de problema de a spune ce tip de matematici, cantitative sau calitative, se pretează cel mai bine sociologiei, nu se pot face decît afirmaţii ipotetice. Este o chestiune pe care viitorul o va decide. Un lucru este însă sigur, şi anume că, în ciuda dezvoltării matematicilor nonmetrice, dacă se face o analiză a ceea ce s-a aplicat pînă în prezent în sociologie, rezultă că ramurile cantitative au dus la rezultatele cele mai importante şi au avut, de departe, cele mai frecvente utilizări. Lucrul este explicabil şi prin legătura strînsă ce există între matematică şi statistică, celălalt instrument fundamental utilizabil pentru ridicarea valorii rezultatelor cercetărilor sociale.

<u>Statistica</u> este o disciplină a cărei natşere, chiar dacă nu poate fi exact datată, se plasează cu siguranţă în secolul XVII. Ea s-a dezvoltat iniţial pe două direcţii, în virtutea celor două puncte de plecare, diferite în spaţiu şi prin problematică, dar aproape simultane în timp. Scoala germană a dezvoltat ceea ce s-ar putea numi "statistica statală", o disciplină chemată să contabilizeze, ordoneze şi să analizeze date prin care poate fi caracterizată dezvoltarea statului într-o etapă sau alta. Scoala engleză - cu reprezentanţi mai cunoscuţi astăzi precum: John Graunt, William Petty sau celebrul astronom Halley - a dezvoltat ceea ce s-a numit iniţial "aritmetica politică", de fapt demografia bazată pe instrumentul statistic. Aparent, şcoala germană, prin obiectivele sale mai largi, ce ofereau o paletă extinsă de aplicaţii pentru noua tehnică, ar fi trebuit să constituie cadrul cel mai favorabil de dezvoltare a statisticii. Lucrurile nu s-au întîmplat aşa deoarece discursurile despre stat şi dezvoltarea lui erau în mod

precumpănitor verbale, aspectul cantitativ-cifric nereuşind să înlocuiască aprecierile impresionist-speculative. In schimb, şcoala engleză, limitată la o arie mai restrînsă de fenomene, a dezvoltat metodologia statistică în punctele sale şi azi esenţiale, fiind, de cele mai multe ori, recunoscută drept singur precursor.

Oricum, nu este în intenţia mea de a face aici istoria statisticii; unicul lucru pe care doresc să-l subliniez cu toată tăria pentru a fi reţinut este faptul că statistica s-a născut şi s-a dezvoltat pe problematica ştiinţelor sociale. Simbioza aceasta a fost atît de puternică încît pînă la sfîrşit secolului trecut, cînd s-au amplificat utilizările ei şi în alte domenii, statistica a fost considerată drept o disciplină socială de sine stătătoare. Deci foarte tîrziu s-a remarcat caracterul metodologic al disciplinei, foarte tîrziu s-a văzut că, de fapt, ea nu are un obiect propriu de studiu şi că metodele sale pot fi aplicate în orice disciplină ştiinţifică, indiferent de domeniul - natural ori social - al obiectului său. Astfel s-a constituit un nucleu ideatic unitar ce poate fi numit "statistica generală" sau "teoria statisticii", format din totalitatea conceptelor, procedurilor şi operaţiilor statistice general valabile. Pe lîngă acestea există desigur concepte şi proceduri specifice aplicaţiilor în anumite ştiinţe particulare, cum sînt economia, demografia etc.

Fără pretenţia de a da o definiţie, se poate spune că statistica este o disciplină ce se distinge prin cel puţin trei elemente esenţiale:

a) caracterul metodologic amintit deja, adică lipsa unui obiect propriu şi faptul că ea constă dintr-un arsenal de tehnici aplicabile în alte discipline;

b) fenomenele, însuşirile etc. studiate cu mijloace statistice se relevă ca aspecte variabile pe o masă mare de unităţi (indivizi);

c) aceste fenomene sînt afectate de o pluralitate de cauze în manifestarea lor şi, ca atare, relaţiile esenţiale se evidenţiază ca o rezultantă a "comportamentelor" individuale.

Aplicaţiile statisticii în domeniul sociologiei sînt multiple mergînd de la oferirea unor modalităţi de analiză a unei însuşiri cu manifestare cantitativă sau calitativă (evidenţierea

distribuţiei indivizilor în diferite clase, calculul procentelor, mediei, abaterii standard etc.) şi pînă la construcţia unor modele sofisticate de detectare a relaţiilor dintre mai multe variabile, oferind astfel argumente obiective pentru încercările de a explica fenomenele sociale. Acest lucru a fost sesizat de la bun început, naşterea statisticii producîndu-se, cum am amintit, tocmai pe tărîmul faptelor sociale. Cele trei secole ce au trecut de la debutul noii perspective de abordare a fenomenelor sociale n-au făcut decît să întărească convingerea în validitatea metodelor statistice, metode care n-au încetat să se amelioreze, să se rafineze, mai ales de la sfîrşitul secolului trecut încoace. Mijloacele electronice de stocare şi prelucrare a informaţiei, apărute în scenă la mijlocul secolului nostru au dat un imbold extraordinar aplicaţiilor statisticii în ştiinţele sociale, iar sociologia a fost unul din beneficiarii care au profitat cel mai mult de noile condiţii.

Relaţiile statisticii cu matematica sînt foarte strînse, nu numai pentru motivul, uşor de detectat şi de înţeles, că ambele lucrează cu cifre. O serie de componente esenţiale ale statisticii şi-au găsit o descriere şi o fundamentare matematică - în speţă în domeniul teoriei probabilităţilor. Intensitatea legăturilor dintre cele două discipline se observă foarte clar în aplicaţii, cînd , deseori, este greu de spus unde se termină statistica şi intră în lucru matematica. Cu toate acestea, ar fi greşit să se considere statistica drept o ramură a matematicii. Distincţia dintre cele două discipline devine evidentă la nivelul discursului în teoria statisticii, unde lipseşte cu desăvîrşite caracterul axiomatic, deductiv chiar, trăsături specifice teoriilor matematice. Apoi, în ciuda faptului că unele noţiuni statistice au un corespondent matematic, ele nu coincid cu aceste imagini ideale ale lor (noţiunea de frecvenţă relativă nu este identică cu cea de probabilitate, curbele de frecvenţe empirice nu reproduc niciodată pe cele ideale etc.). Mai mult, o serie de noţiuni statistice - indicatori, indici, coeficienţi etc. nu au nici o justificare matematică precisă, ei fiind rezultatul exclusiv al demersului statistic, al analizei datelor empirice. De aceea nu este de mirare că pentru a măsura acelaşi lucru întîlnim, în statistică, trei, patru sau chiar mai multe mărimi propuse; o asemenea "dezordine" nu este

evident, posibilă în matematică.

O imagine asupra unor instrumente de analiză statistică va fi oferită în ultimele două părţi ale cursului de faţă. In încheierea acestui capitol, voi preciza doar cîteva noţiuni de bază, ele fiind, de fapt, rezultatul întregii discuţii de pînă aici. Aşa cum am spus, statistica lucrează cu mase de indivizi, plecînd de la descrierea modului în care variază diferitele fenomene în rîndul acestora. Aceasta înseamnă - după terminologia folosită în primul paragraf - că statistica, deşi are ca obiect de analiză însuşirea, utilizează exclusiv perspectiva cantitativistă la nivelul obiectului-suport.

Drept consecinţă, noţiunile primare ale statisticii sînt cele de individ sau, echivalent, unitatea statistică, reprezentînd obiectele-suport purtătoare de însuşiri şi de univers (sau echivalent: populaţie, colectivitate statistică), în calitate de mulţime a tuturor obiectelor pe care se studiază o însuşire dată. Termenul de individ - şi, corelat, cel de populaţie - nu trebuie să ne inducă în eroare. Nu e vorba de (sau numai de) indivizi umani; această terminologie va fi aplicată oricărei unităţi de analiză: familii, grupuri, oraşe, sate, instituţii, naţiuni etc. Individul statistic este deci ceea ce se numără, iar însuşirea - sau caracteristica, în limbaj statistic consacrat - este cea care se evaluează cu o anumită scală. Se vor distinge, prin urmare, metode statistice calitative sau nonparametrice, dacă scala este nominală sau ordinală, şi metode statistice cantitative, cînd ea e de intervale sau rapoarte, deci cînd se realizează măsurarea însuşirii. Aşadar, atributul de cantitativ aplicabil unor proceduri statistice vizează în exclusivitate trăsătura, deoarece la nivel de obiect abordările statistice sînt prin definiţie cantitative.

Ca şi în cazul matematicii, folosirea mijloacelor de analiză statistică în sociologie - făcută cu discernămînt şi fără exces - nu poate avea decît efecte pozitive asupra limbajului, raţionamentului, şi în general, asupra întregului demers descriptiv-explicativ al acestei ştiinţe.

PARTEA II

METODE DE CULEGERE A INFORMAŢIEI

Cap.IV. ANCHETA

Fără teama de a greşi, se poate aprecia că ancheta este metoda de cercetare cu cea mai largă utilizare în sociologia empirică. Această frecvenţă ridicată de apariţie a metodei în studiile "de teren" n-a întîrziat să creeze o imagine falsă în ochii nespecialiştilor, care o concept ca metodă exclusivă a sociologiei, iar sociologul este adesea reprezentat, într-o imagine populară, drept "omul cu chestionarul". Indiferent de criticile destul de severe adresate de-a lungul timpului acestei modalităţi de culegere a informaţiei, ancheta a rămas şi astăzi o metodă de bază, iar literatura metodologică referitoare la ea a proliferat atît de mult încît orice încercare, ca cea de faţă, de a concentra în cîteva pagini ideile principale legate de utilizarea sa reprezintă, pur şi simplu, un act de temeritate şi, desigur, implică riscul de a omite aspecte deosebit de importante.

Ancheta, ca metodă specifică sociologiei, a apărut din particularitatea pe care unele dintre "obiectele materiale" de cercetare (oamenii) o au, şi anume aceea de a permite un schimb direct de informaţie între cercetător şi "obiect", prin intermediul limbajului. Renunţarea la ancheta ar însemna, în ultimă instanţă, renunţarea la acest avantaj al comunicării interumane, deci la privarea sociologiei de o sursă de informaţii, care, chiar dacă ridică, de cele mai multe ori, probleme metodologico-tehnice şi interpretative foarte dificile, îi poate aduce un inedit beneficiu cognitiv.

Perfecţionările metodice realizate pînă în prezent - şi posibil de continuat în viitor - au condus la creşterea gradului de validitate a cunoştinţelor dobîndite prin intermediul anchetei, astfel încît deja se poate vorbi realmente de ea ca de o modalitate fundamentală în studierea, faptelor, fenomenelor şi chiar a proceselor sociale. Din păcate, publicul larg cunoaşte foarte puţine

din dificultățile ridicate de utilizarea efectivă a metodei și nu rareori ea este discreditată de încercări ale amatorilor, care văd aici o posibilitate extrem de facilă de a întreprinde ceva capabil să le eticheteze activitatea cu atributul, atît de rîvnit, de "științifică". De aceea, în capitolul de față, voi încerca - din cauza întinderii neobișnuite a problematicii - să abordez doar unele aspecte, capabile să creeze, sper, o imagine de ansamblu a chestiunilor care se ridică, să atragă atenția asupra dificultăților metodico-practice și, evident, să convingă pe cel care ar intenționa să activeze pe tărîmul cercetării sociale să se îndrepte mai întîi spre lectura unor lucrări.

1. Metoda și tehnicile sale

In primul capitol al cursului am prezentat o clasificare a metodelor sociologiei, în rîndul cărora am inclus ancheta și interviul, ca metode distincte, sugerînd în treacăt și deosebirea dintre ele. Este locul să reluăm problema, mai cu seamă că forma de clasificare pe care am utilizat-o impune de la sine o definire a fiecărei metode numai prin raportare la specificul celeilalte.

Fără pretenția unor definitii riguroase, vom admite că ancheta este metoda de culegere a informației de la o masă mare de indivizi umani, cu ajutorul unui instrument specific: chestionarul. La rîndul său, chestionarul poate fi înțeles ca o mulțime de întrebări, de regulă scrise, adresate subiecților în forme variate, într-o anumită succesiune, pe baza unor considerente logice și psihologice.

Din punct de vedere tehnic, de realizare practică efectivă, metoda anchetei are două variante fundamentale;

a) ancehta directă sau orală și

b) ancheta indirectă sau în scris.

In primul caz, întrebările chestionarului se adresează subiecților anchetați sub formă orală, de către cercetător sau de persoane auxiliare folosite de acesta doar pentru culegerea informației, numite operatori de anchetă. In al doilea, sociologul ia doar un contact indirect cu subiectul, întrucît acesta din urmă este chemat să citească și să completeze singur chestionarul-formular care i se

înmînează într-o formă sau alta.

Tehnica anchetei directe se realizează, de cele mai multe ori, print-un dialog față în față între cercetător (operator) și subiectul supus investigației (sau, cum se spune în franceză, prin "entretien personnel"), dar, în anumite cazuri, cu totul deosebite, ea poate lua și forma unui dialog prin telefon. Dat fiind faptul că ancheta prin telefon este foarte rar folosită, nu voi mai reveni în continuare asupra ei, menționînd aici că ea se realizează doar pentru colectarea unei cantități reduse de informație și de la un public nu prea numeros. Mai mult, aria problematicii astfel abordate este restrînsă, deoarece nici în cele mai dezvoltate țări telefonul nu se găsește în fiecare casă; deci populația vizată de cercetare trebuie să fie inclusă în populația care posedă acest mijloc de comunicare.

Ancheta indirectă se poate efectua în mai multe forme de realizare concretă. Una dintre cele mai vechi, deci dintre primele utilizate, constă în publicarea chestionarului în ziare și reviste și colectarea răspunsurilor tuturor celor care doresc să returneze formularul completat. Astăzi acest procedeu se mai practică doar în scopuri publiciste și publicitare, cercetările științifice serioase renunțînd de mult la el, pentru motivul principal că, oricît de mare ar fi numărul chestionarelor returnate - și el ajunge adesea la ordinul zecilor de mii -, persoanele care răspund la asemenea solicitări nu se constituie într-un eșantion cu o reprezentativitate suficient de ridicată spre a putea generaliza, cît de cît corect, constatările asupra întregii populații.

O altă formă de materializare a tehnicii anchetei indirecte constă în strîngerea unei mulțimi de persoane într-o sală în scopul completării chestionarului. De regulă, operația se realizează sub controlul cercetătorului, care distribuie chestionarele, explică scopul anchetei și precizează modalitatea de completare. Intervenția cercetătorului poate lua forme variate, mergînd de la cîteva indicații tehnice prealabile și supravegherea subiecților pentru a nu "colabora" la completare pînă la citirea fiecărei întrebări în parte și chiar explicarea sau comentarea ei. Pe această cale se poate deci ajunge atît de departe încît să se iasă efectiv din cadrul tehnicii de anchetă indirectă propriu-zisă, fie în direcția unei anchete orale sui-generis, fie spre o tehnică psihoso-

ciologică consacrată sub numele de interviu de grup. Bineînţeles că forma de anchetă colectivă se poate realiza doar în cazul cînd sînt îndeplinite unele condiţii speciale. Pe de o parte, este nevoie ca procedura de eşantionare să fie adecvată acestui scop (eşantionare multistadială, de regulă), iar, pe de alta, să avem efectiv posibilitatea să reunim mai multe persoane într-o sală, ceea ce este mai simplu atunci cînd anchetăm oamenii la locul de muncă, în şcoli, facultăţi etc.

A treia specie a anchetei indirecte se realizează prin înmînarea chestionarului, în mod individual, fiecărui subiect de către un membru al echipei de cercetare sau auxiliar. Si aici întîlnim un evantai larg de variante, ce corespund în mare celor din cazul precedent. Adică, la un pol se poate imagina situaţia în care cercetătorul realizează doar un contact superficial cu subiectul atît cît să-i fie necesar să-l determine la completarea chestionarul şi să-l lămurească asupra aspectelor tehnice ale înregistrării răspunsurilor, iar la altul cînd ancheta indirectă abia mai poate fi distinsă de cea directă, într-atît de implicat devine cercetătorul în operaţia de completare a formularelor. In situaţiile apropiate de forma mai impersonală, cele mai des întîlnite, chestionarul - după lămuririle prealabile - poate fi lăsat în posesia persoanei chestionate mai multe zile, urmînd a se reveni pentru ridicarea lui.

In sfîrşit, o a patra modalitate de anchetă indirectă este cea realizată prin poştă: chestionarele se trimit subiecţilor ce formează eşantionul, însoţite de o scrisoare în care se precizează scopul anchetei, felul cum s-au ales persoanele anchetate şi se dau instrucţiunile de completare. De asemenea, prin cele mai politicoase şi persuasive formule, se va solicita returnarea chestionarului completat, în care scop se va ataşa un plic timbrat cu adresa instituţiei ce efectuează cercetarea.

Din sumara trecere în revistă a diverselor tehnici de anchetă, cred că se poate desprinde specificul metodei, fiind mai uşor acum de precizat diferenţa faţă de interviu. In esenţă, ancheta, indiferent de modalitatea concretă de realizare, presupune, aşa cum am precizat, apelul la un chestionar, adică la un set de întrebări relativ clar structurate, ce se adresează, direct sau indirect, subiecţilor, într-o formă standard, sau în orice caz, apropia-

tă de aceasta. Interviul, chiar dacă face apel la un instrument similar chestionarului, numit ghid de interviu, se deosebeşte de anchetă prin aceea că cercetătorul realizează în fapt o discuţie relativ liberă cu subiectul, ghidul de interviu avînd drept unic scop să-l ajute pe cel care conduce convorbirea să o menţină în cadrele iniţial fixate. Se vede deci că interviul poate fi apropiat de ancheta directă şi nu puţine manuale de metode şi tehnici le tratează indistinct. Mai exact, se vorbeşte de o aceeaşi metodă care ar lua forme diverse, mergînd pe o anumită scală de formalizare de la discuţia absolut liberă pînă la cea perfect structurată , pe baza unui chestionar care nu permite nici o înregistrare necodificată a informaţiei furnizate de subiect şi nici o libertate pentru anchetator în a schimba textul întrebărilor şi ordinea lor.

Este adevărat că cele două metode interferează, că în unele situaţii ancheta poate lua turnură de interviu, dar cred că mai raţională este distingerea lor aşa cum am făcut-o în paginile de faţă. Pe lîngă instrument, un alt factor care pledează pentru o atare distincţie este şi numărul subiecţilor investigaţi. Intr-adevăr, în cazul interviurilor acest număr nu poate fi decît mic sau foarte mic, deoarece, pe de o parte, durata unui interviu este net superioară celei a unei anchete, apoi, pe de altă parte, numărul persoanelor capabile să realizeze interviuri este inferior celui al operatorilor ce pot fi antrenaţi în anchete, deoarece pentru aceasta din urmă formă nu e nevoie de o calificare deosebită (practic interviul nu poate fi făcut cu personal auxiliar) şi , în fine, prelucrarea datelor unui număr mare de interviuri este, dacă nu imposibilă, extrem de greu de realizat, avînd în vedere caracterul absolut nestandardizat al formelor de înregistrare: text liber, scris ori stenografiat, sau bandă de magnetofon. Ancheta, dimpotrivă, tocmai datorită chestionarului, poate fi efectuată pe un eşantion mare de persoane, căci implică riscuri mult mai reduse în folosirea anchetatorilor nespecialişti şi oferă posibilitatea unor prelucrări rapide ale datelor, mai ales acum în era calculatoarelor electronice.

O consecinţă firească a modului de realizare constă în faptul că interviul permite abordarea problemelor în profunzime,

discuţia putînd fi purtată asupra celor mai mici detalii sau a celor mai subtile aspecte psihologice. Ancheta este, prin natura lucrurilor, condamnată să se mărginească la aspecte mai simple, mai uşor de investigat, la probleme ce pot fi abordate într-o discuţie rapidă sau pe care oamenii sînt dispuşi să le clarifice în cadrul unui formular ce le este pus în faţă spre completare. De aici rezultă deci avantajele şi dezavantajele fiecărei metode: ancheta este superioară datorită numărului mare de indivizi chestionaţi, ceea ce asigură o reprezentativitate ridicată eşantionului, deci garanţia că datele obţinute sînt relevante pentru întreaga populaţie; interviul vine cu argumentul calităţii şi al profunzimii, cu o mai largă deschidere spre tema studiată. In aceste condiţii mi se pare natural ca, aşa cum am mai spus-o , cele două tehnici să se completeze reciproc. Forma optimă de împletire a lor, recomandată în unele lucrări de specialitate, ar fi următoarea. Cercetarea ar trebui să înceapă cu un interviu al unui lot restrîns de subiecţi, ceea ce va permite cercetătorului familiarizarea cu problema studiată, descoperirea unor aspecte, dimensiuni nebănuite ale fenomenelor abordate, schiţarea unor posibile scheme explicative sau îmbogăţirea arsenalului de ipoteze etc. A doua fază ar constitui-o ancheta pe un lot mare şi cu instrumente standardizate, a căror construcţie este evident facilitată de interviurile de etapa precedentă. In fine, pornind de la concluziile degajate din anchetă, se recomandă reluarea interviurilor (pe acelaşi lot ca şi la început sau pe altul), de data aceasta cu scopuri mult mai clar conturate, pentru a întregi constatările generale, pentru a găsi explicaţii adecvate unor corelaţii formale etc.

După această scurtă paranteză, să revenim la tehnicile de anchetă, să vedem ce probleme practice ridică fiecare, ce avantaje sau dezavantaje însoţesc alegerea unei anumite tehnici într-o cercetare sociologică efectivă.

Mai întîi o remarcă: manualele de metode şi tehnici par să încline în a acorda un avantaj decisiv în formarea anchetelor directe (orale), avantaj ce ar rezulta, aşa cum foarte concis se exprimă Caplow, din faptul că oamenii găsesc mult mai multă plăcere în a purta o conversaţie decît în a completa un formular (1970, p.222). Ideea aceasta,în forme mai mult sau mai puţin sofis-

ticate, se regăsește în bună parte din lucrările celor care au scris despre anchete. De fapt, se poate întîlni uneori și o notă de surpriză legată de succesul pe care-l au operatorii de anchetă pe lîngă persoanele chestionate. Dacă cei dintîi sînt bine pregătiți, au o atitudine și un comportament adecvate față de subiecți, ei pot ajunge la a aborda cele mai delicate probleme din viața intimă a indivizilor, așa cum au demonstrat-o anchetele legate de comportamentul sexual (Kinsey).

Madeleine Grawitz (1972, p.653) consideră că, în general, succesul anchetei ar putea fi explicat pe baza a trei factori principali, dintre care ultimul joacă efectiv în favoarea celei directe:

a) Reflexul de politețe. Este greu să refuzi să-i vorbești cuiva care ți se adresează politicos, te roagă să-l ajuți în realizarea muncii ce-o are de îndeplinit etc.

b) Dorința de influențare. Cu cît subiectului i se sădește o comvingere mai puternică despre faptul că prin răspunsurile sale va putea contribui la schimbarea unor situații, la ameliorarea unor stări de lucruri care-l privesc direct, el va fi dispus într-o măsură mai mare să se angajeze în discuție.

c) Nevoia de a vorbi. E vorba nu de sensul vulgar de pălăvrăgeală, ci de unul mai adînc, psihologic, de nevoia de comunicare și de a fi înțeles. Dacă n-ar exista aceste nevoi, oricîtă politețe și bunăvoință ar manifesta subiectul, ar fi totuși de neconceput ca o persoană singură acasă (mai ales în mediul urban) să deschidă ușa unui necunoscut spre a-i răspunde la întrebări, trecînd peste toate riscurilor pe care un astfel de gest le poate implica.

Așadar, tehnica anchetei directe se dovedește a fi mai atractivă pentru subiecți, de unde rezultă unul din avantajele sale: ponderea nonrăspunsurilor este, în general, mai redusă decît în ancheta indirectă. Intr-adevăr , aceasta din urmă , mai ales dacă este vorba de chestionarul prin poștă, ridică nenumărate obstacole în calea unei interpretări corecte a rezultatelor tocmai din cauza numărului mare de persoane care nu returnează chestionarul. In felul acesta, rezultatele anchetei sînt, în principiu, distorsionate dat fiind că, foarte probabil , persoanele care nu răs-

spund au alte caracteristici decît cele care răspund. In consecință nu va fi suficientă doar o supradimensionare a eșantionului inițial în speranța că prin nonrăspunsuri se va obține eșantionul de mărime dorită; oricît de mulți indivizi am include, gradul de distorsiune nu se reduce, căci el depinde de proporția nonrăspunsurilor și nu de numărul acestora.

Ideea de mai sus poate primi o formulare matematică foarte clară. Să notăm cu p proporția persoanelor din eșantion care răspund la chestionar și cu $1-p$ a celor care nu răspund. Să mai notăm cu x proporția indivizilor din eșantionul total care posedă o anumită caracteristică, cu x_1 proporția corespunzătoare în rîndul celor care răspund și cu x_2 în subeșantionul celor care refuză să răspundă. Atunci proporția x se poate scrie:

$$x = px_1 + (1-p)x_2$$

Efectuînd cercetarea și obținînd numai o parte din chestionare, vom ajunge să determinăm pe x_1 în loc de x, sau, altfel spus, se va produce o distorsiune d egală cu diferența dintre cele două valori:

$$d = x - x_1$$

Inlocuind pe x din formula precedentă, obținem:

$$d = px_1 + (1-p)x_2 - x_1 = (1-p)x_2 - (1-p)x_1$$

sau

$$d = (1-p)(x_2 - x_1)$$

Prin urmare, distorsiunea introdusă este direct proporțională cu 1-p, adică cu ponderea nonrăspunsurilor, și cu $x_2 - x_1$, deci cu diferența de comportament între indivizii care răspund și cei care nu răspund. Distorsiunea se va anula atunci cînd unul dintre factori este zero. Or, cum, de regulă, despre al doilea factor noi nu avem suficiente cunoștințe, nu vom putea presupune decît cu multe riscuri că el este nul. Rămîne, așadar, ca mult mai sigură calea de anulare a primului factor, adică străduința de a elimina pe cît posibil nonrăspunsul.

Calculul schiţat mai sus este, evident, valabil pentru orice gen de anchetă, ceea ce susţine afirmaţia că problema nonrăspunsului este una generală. Numai că gravitatea ei diferă în funcţie de diversele forme pe care ancheta le poate lua. Altfel spus, nonrăspunsul apare în toate anchetele, dar proporţiile în care el se prezintă variază foarte mult de la un tip de anchetă la altul, minimul fiind înregistrat în cea directă iar maximul în cea indirectă prin poştă.

Un anumit procent de nonrăspunsuri nu poate deci fi eliminat în nici o anchetă, dar dacă valoarea sa este mică, atunci distorsiunea produsă este şi ea scăzută. Să exemplificăm puţin lucrurile. Să presupunem că trăsătura urmărită apare la 7o% dintre indivizii care răspund şi doar la 5o% din cei care nu răspund, adică cu notaţiile de mai sus:

$$x_1 = o,7o$$

$$x_2 = o,5o$$

Dacă proporţia nonrăspunsurilor este de 1o% (1-p=o,1o), cifră ce caracterizează unele anchete directe, atunci distorsiunea va fi:

$$d = o,1o.o,2o = o,o2$$

deci o valoare de numai 2 procente, ce se încadrează, de regulă, în marja de eroare acceptabilă în astfel de cercetări. Dimpotrivă, dacă cei care nu răspund sînt în proporţie de 8o% (1-p=o,8o), aşa cum se întîmplă la chestionarul prin poştă, distorsiunea va primi valoarea:

$$d = o,8o.o,2o = o,16$$

adică 16 procente, ceea ce va afecta serios concluziile obţinute.

Un al doilea avantaj al anchetei directe îl constituie faptul că, de cele mai multe ori, în asemenea situaţii persoanele interogate acordă o atenţie mai mare chestionarului. Faptul că răspunsul este dat unei persoane aflată în faţa ta are desigur o anumită influenţă asupra grijei de a înţelege corect fiecare întrebare, de a nu te contrazice, şi în general, de a privi cu seriozitate întreprinderea respectivă. Avînd doar chestionarul în faţă

eşti tentat să-l completezi cît mai repede pentru a te scăpa de o grijă, de unde o anumită neglijenţă, superficialitate etc. Desigur, lucrurile nu stau întotdeauna astfel, căci există anumite categorii de persoane care sînt mai conştiincioase în completare decît operatorii înşişi, dar ele nu reprezintă decît excepţiile care confirmă regula.

Un al treilea avantaj ce poate fi invocat pentru ancheta directă este spontaneitatea cu care se dau răspunsurile la chestionar în acest caz. Problema spontaneităţii este relevantă în cazul întrebărilor la care se cer opiniile, părerile, atitudinile oamenilor şi nu atît anumite fapte. Intr-adevăr, pentru primul gen de întrebări este deosebit de important ca răspunsurile să fie provocate cît mai direct posibil. O perioadă de meditaţie, facilitată de posesia chestionarului, conduce, de regulă, la alterarea opiniilor respective, fie în sensul că omul are tendinţa de a se consulta cu semenii săi apropiaţi (şi atunci se amestecă opiniile mai multora), fie că va medita asupra eventualelor consecinţe ce le-ar putea avea pentru el emiterea unor păreri şi va tinde să se conformeze unor opinii dezirabile din punct de vedere social. De altfel, chiar şi fără răgaz de gîndire opinii false (spre deosebire de fapte false) sînt mai uşor de exprimat în scris decît în faţa cuiva, cel puţin din motive de jenă.

In fine, anchetele indirecte sînt posibile numai în cazurile în care nu este nici un pericol de contaminare a răspunsurilor la diferite întrebări, contaminare ce se produce atunci cînd subiectul are posibilitatea să citească dinainte toate întrebările chestionarului. De regulă, în chestionar - mai cu seamă la întrebările de opinie - are importanţă şi ordinea în care acestea sînt adresate, recurgîndu-se la o serie de stratageme pentru a obţine şi pe această cale răspunsuri cît mai corecte. Or, în ancheta indirectă astfel de artificii sînt imposibile.

Cu siguranţă că şi pentru anchetele indirecte se pot pune în evidenţă o serie de avantaje, altfel ele ar fi dispărut din practica cercetării sau nici n-ar fi fost inventate. In primul şi în primul rînd aceste valaneţe sînt de ordin material. Costul cercetării în anchetele indirecte - luat termenul de cost într-un sens larg ce implică nu numai mijloace băneşti dar şi umane - este mult redus, căci fie se elimină complet deplasarea operatorului la su-

biect (chestionarul prin poştă) , fie un cercetător anchetează simultan un număr mare de subiecţi (anchete colective cu autocompletarea chestionarului), fie timpul pierdut chiar cu aplicarea individuală este redus la minimum.

In al doilea rînd, nu este imposibil să se imagineze şi situaţii în care ancheta indirectă aduce şi avantaje de ordinul cunoaşterii. Intr-adevăr, dacă informaţia cerută este de natură factuală, ea poate presupune un anumit timp de gîndire sau chiar o activitate prealabilă de documentare din partea subiectului, lucruri ce pot fi mai lesne făcute dacă operatorul nu este de faţă. Dar trebuie s-o recunoaştem astfel de situaţii sînt mai rare şi, din punctul de vedere al acurateţei datelor, este preferabil de cele mai multe ori ancheta directă. Afirmaţia devine şi mai convingătoare dacă ne gîndim că, în ciuda eliminării analfabetismului, un procent nonneglijabil din populaţie nu are deprinderi deosebite de scris şi de citit. A pune în faţa unor asemenea oameni un chestionar cu 2-300 întrebări nu înseamnă numai a-i expune unei activităţi dificile, ci şi a te expune pe tine ca cercetător riscului ca la o parte din întrebări să nu fie înregistrate corect răspunsurile, ca o altă parte să nu fie corect înţelese sau ca, din cauza efortului, de la un moment dat subiectul să treacă pur şi simplu prin chestionar fără nici un interes sau chiar să abandoneze completarea lui.

Din toate cele de mai sus, cred că rezultă cu claritate că ancheta indirectă poate fi utilizată atunci cînd avantajele materiale impun acest lucru, dar numai cu condiţia ca eventualele pierderi de ordin cognitiv să fie minime. In general, chiat dacă tipul de informaţie solicitat şi chestionarul întocmit nu ridică probleme în ceea ce priveşte autocompletarea, este recomandabil să se recurgă la această tehnică numai în cazul în care populaţia vizată de cercetare posedă în întregime un anumit nivel de instrucţie (e preferabil să aibă cel puţin liceul terminat). In caz contrar, riscurile de erori şi nonrăspunsuri sînt atît de mari încît economiile realizate nu pot justifica adoptarea tehnicii anchetei indirecte.

Datorită faptului că în anchetele directe un rol important revine operatorilor, nu numai ca elemente intermediare, mijlocitoare în procesul de circulaţie a informaţiei de la subiectul investigat la cercetător, deci nu numai ca un, să zicem, conductor

pasiv, ci, în primul rînd, ca elemente active care provoacă emiterea informaţiei, mă voi opri, în cîteva cuvinte, asupra activităţii lor şi a calităţilor pe care trebuie să le îndeplinească. Sublinierea, oricît de puternică, a importanţei activităţii operatorilor de anchetă nu este niciodată de prisos. Parafrazînd pe Moser, care spune că o anchetă nu poate fi mai bună decît chestionarul său - ceea ce este perfect adevărat - putem formula tot atît de îndreptăţit şi ideea că o anchetă nu este mai bună decît lotul operatorilor folosiţi. In adevăr, fără a exagera cu nimic lucrurile, valoarea întregii investigaţii depinde pînă la urmă de capacitatea acestor persoane de a convinge subiecţii să poarte discuţia şi să o poarte în mod cît mai sincer, de conştiinciozitatea lor în respectarea tuturor indicaţiilor tehnice de desfăşurare a anchetei, de rigoarea cu care înregistrează răspunsurile primite etc.

Marile institute de opinie publică din occident dispun de o reţea de operatori de anchetă cu care sînt capabile să acopere întreg teritoriul naţional şi să realizeze investigaţii într-un timp foarte scurt, pe baza unor comenzi şi referitoare la evenimente cu derulare rapidă. Operatorii pot fi uşor mobilizaţi în astfel de campanii, deşi ei nu practică o asemenea muncă decît într-un mod ocazional, nefiind deci vorba de o profesie. De altfel, reţeaua operatorilor este mereu reîmprospătată, considerîndu-se că o vechime prea mare în această activitate impietează asupra rezultatelor.

La noi în ţară, actualmente, practic, nici o instituţie nu facel apel la asemenea operatori plătiţi. Se utilizează, pentru completarea chestionarului, diferite persoane care pot fi mobilizate, de o manieră sau alta, la acest gen de acţiuni: studenţi, cadre didactice etc. De aici rezultă necesitatea acordării unei griji deosebite cel puţin pregătirii şi controlului muncii unor astfel de operatori, dacă nu este posibilă şi în selecţionarea lor. Tot aşa se poate explica şi proporţia relativ mare pe care o deţin la noi anchetele indirecte, cu toate dezavantajele pe care acestea le prezintă; dar mulţi cercetători sau colective de cercetare nu au efectiv posibilitatea de a utiliza anchetatori.

Se discută în literatura de specialitate destul de mult asupra calităţilor pe care ar trebui să le posede o persoană spre a fi un bun operator- Deşi este evident că anumite trăsături psi-

hologice şi chiar o înfăţişare fizică plăcută pot avea influenţe notabile asupra succesului muncii sale, pînă la urmă calitatea fundamentală a operatorului trebuie să fie totuşi o trăsătură de caracter, şi anume conştiinciozitatea. Orice om normal, deci neafectat de anumite defecte fizice sau intelectuale evidente, care posedă un înalt grad de conştiinciozitate, urmărind să-şi ducă la bun sfîrşit munca, respectînd toate instrucţiunile primite, poate fi considerat un operator bun. Dacă această trăsătură de caracter lipseşte, ea nu poate fi suplinită prin nici o altă calitate. Restul însuşirilor fizice sau de personalitate pot doar influenţa pozitiv sau negativ, pot ridica sau scădea, într-o anumită măsură, nivelul calitativ al muncii. Mai mult, prin antrenament sau efort de voinţă, unele dintre trăsăturile de personalitate cu efecte negative asupra anchetei, dacă sînt conştientizate, pot fi serios modificate, cel puţin pe durata acestei munci.

Se cere, de obicei, ca operatorul să fie o persoană deschisă, extrovertită, cu facilităţi de comunicare cu ceilalţi. Ea trebuie să fie capabilă să trezeazscă interesul subiectului pentru anchetă şi, prin urmare, să posede o capacitate de abordare a oamenilor de o manieră plăcută, destinsă etc. Prin felul său de a se purta , la început dar şi pe parcursul anchetei, operatorul trebuie să se facă simparizat de către subiect, ba chiar să-i insufle încredere. Se cere deci ca el să fie capabil să poarte discuţii pe teme diverse, să fie intereat de problemele generale ale semeneilor săi, să-şi poată comvinge interlocutorul că-l înţelege în durerile şi necazurile pe care acesta le are. Desigur, asemenea însuşiri, dacă se manifestă prea puternic, pot avea un efect contrar celui dorit. Implicarea operatorului în problemele personale ale anchetatului poate contribui la influenţarea lui, la crearea unei deformări constante în aprecierea informaţiei solicitate prin chestionar.

Este evident că operatorului i se cere un anumit grad de inteligenţă, care-l va ajuta să se descurce în cele mai neaşteptate situaţii, permiţîndu-i adaptarea în stabilirea relaţiilor cu foarte multe categorii de persoane întîlnite, diferite între ele nu numai prin nivelul cultural dar şi printr-o serie de trăsături psihologice şi de comportament. Nu rareori el este ţinta unor atitudini agresive, de la ironiile fine ale unui intelectual pînă la

comportamentul plin de vulgarităţi al unui beţiv şi intrînd în casă, el trebuie să ajungă să comverseze cu toţi aceşti oameni, să ştie să-i găsească fiecăruia "coarda sensibilă" prin care să-l atragă spre discutarea problemelor urmărite.

Se mai apreciază că o calitate necesară a oricărui bun operator de anchetă este exactitatea, rigoarea. Intr-adevăr, un minim spirit de rigoare nu trebuie să lipsească în munca nici unui anchetator, deoarece altfel este de aşteptat apariţia unui număr exagerat de mare de erori în înregistrarea informaţiei. Dar consider că această însuşire - ca şi cele menţionate mai sus - poate fi dobîndită printr-un antrenament adecvat şi aproape oricine poate ajunge să completeze exact un chestionar.

Referitor la nivelul de instrucţie al operatorilor, specialiştii înclină spre a-i considera mai potriviţi pe cei cu unul mediu, dat fiind că persoanele cu un standard cultural foarte ridicat ajung să se plictisească mai repede de operaţiile, la urma urmei de rutină, pe care le presupune completarea unui număr mare de chestionare. Desigur, această apreciere poate fi discutată; dacă operatorul manifestă un interes oarecare în cercetare, dacă el este stimulat să participe la investigaţie şi din alt punct de vedere decît ca simplu operator, atunci nivelul de cultură şi chiar unele cunoştinţe de specialitate pot să fie de bun augur pentru cercetare. Am ajuns astfel să ating şi această problemă delicată a interesului operatorului pentru cercetare. Ea este realmente o problemă, mai ales la noi unde, cum spuneam, munca operatorilor nu este remunerată; fără un minim de interes din partea celor angajaţi la realizarea anchetei, nu este de conceput o cercetare reuşită. Fie că este vorba de studenţi, de cadre didactice sau de alte persoane folosite ca operatori, ele trebuie, pe de o parte, să înţeleagă importanţa, ştiinţifică sau practică, a studiului şi, pe de alta, să fie implicate în cît mai mare măsură şi în alte faze ale cercetării.

Rezultă deci, din cele spuse pînă aici, că, şi în cazurile în care sociologul nu are prea multe variante de alegere a echipei de operatori, se cer cîteva lucruri intreprinse pentru a nu ajunge la un eşec complet al cercetării.

In primul rînd, trebuie eliminate din lotul potenţial de operatori acele persoane despre care se ştie de la bun început

că au grave carenţe de ordin moral, fizic sau de altă natură, persoane care prin prezenţa lor în cîmpul cercetării ar putea constitui o indubitabilă sursă de erori.

In al doilea rînd, să se asigure o instruire completă şi complexă a operatorilor cu care se va lucra, prin care să se urmărească obiectivele următoare:

(i) Cunoaşterea exactă a scopului şi importanţei cercetării. La acest capitol se vor accentua efectele şi beneficiile pentru societate sau pentru dezvoltarea sociologiei, contribuind astfel la stimularea motivaţiei participării operatorilor la buna desfăşurare a anchetei. Se va insista asupra răspunderii ce revine întregii echipe de cercetare pentru obţinerea unei informaţii corecte, nedistorsionate, lucru uşor de realizat atunci cînd rezultatele finale vor fi puse la dispoziţia organelor de conducere, constituind o bază pentru luarea deciziilor. De asemenea, se pot preciza, aşa cum am mai spus, eventualele posibilităţi de participare a celor doritori la fazele de prelucrare şi interpretare a datelor.

(ii) Familiarizarea operatorilor cu chestionarul ce urmează a fi aplicat. Fiecare dintre aceştia trebuie să cunoască sensul exact al tuturor întrebărilor, felul în care acestea se adresează, precum şi modalitatea concretă de înregistrare a datelor. Este recomandabil ca în prezenţa întregului lot de operatori să se realizeze completarea cîtorva chestionare, în situaţii de simulare. (Unul dintre sociologii cu ma multă experienţă să joace rolul subiectului imaginar iar cîţiva dintre operatori să încerce pe rînd să-l chestioneze). Se vor evidenţia cu această ocazie toate posibilităţile permise operatorilor în a schimba eventual ordinea întrebărilor în a da explicaţii suplimentare atunci cînd o întrebare nu este corect înţeleasă (eventualele libertăţi în reformularea textului întrebărilor) etc.

(iii) Precizarea modalităţilor de găsire şi identificare a persoanelor de chestionat şi alternativele pentru cazul cînd persoanele în cauză nu pot fi găsite, refuză sau, din alte motive, se află în imposibilitatea de a răspunde la întrebări. Se vor transmite cu această ocazie informaţii asupra modului de alegere a eşantio-

nului, cît şi unele sumare cunoştinţe legate de noţiunea de reprezentativitatea a acestuia. Astfel, operatorii vor fi puşi în gardă că orice abatere de la regula de alegere a indivizilor stabilită constituie o sursă de eroare, care poate anula multe din eforturile lor de a conduce cît mai exact ancheta.

(iv) Specificarea posibilităţilor concrete de rezolvare a tuturor problemelor ce se pot ivi pe teren, legate de aplicarea chestionarului şi, în general, de activitatea lor de investigatori. Se va atrage cu deosebire atenţia că toate problemele ce apar trebuie rezolvate de o manieră uniformă, în acelaşi mod pentru toţi operatorii, şi că inovaţiile personale ale unuia sau altuia nu vor avea alt efect decît să conducă la inutilizarea chestionarelor sau, în cel mai bun caz, a unor părţi din chestionarele astfel modificate.

In al treilea rînd, este necesar să se asigure un control cît mai riguros posibil al muncii de teren. Operatorii nu pot fi, decît în cazuri cu totul excepţionale, distribuiţi direct în teren şi reîntîlniţi doar în finalul cercetării, ci trebuie urmăriţi pas cu pas în activitatea lor. E vorba, desigur, mai ales de a verifica primele chestionare efectuate de fiecare, spre a se elimina din timp greşelile inerente ce apar din cauza neînţelegerii corecte a sensului unor întrebări, a modalităţilor de înregistrare a informaţiei sau datorită altor factori. Apoi, dacă lucrul este posibil, se cere verificat, prin sondaj, dacă adresele primite de fiecare au fost într-adevăr respectate şi eventual chiar repetarea unor interviuri de către sociolog. Este limpede că sarcinile controlului se vor reduce substanţial dacă recomandările precedente au fost cît mai riguros realizate. De asemenea, este evident că orice operator surprins că se pretează la fraudă trebuie eliminat din echipă; la fel şi cei care, din diferite alte motive, aduc chestionare cu un număr nepermis de mare de erori.

Avînd în vedere că, în multe cazuri, lotul operatorilor este constituit din persoane care-şi oferă voluntar, sau aproape, serviciile se cere ca întreaga activitate de control să se facă de o manieră care să nu lezeze personalitatea acestora, să nu jignească pe nimeni. Asta nu trebuie să ducă însă la trecerea cu vederea a nici unei neregului. E nevoie deci de o îmbinare optimă a fer-

mităţii cu spiritul de politeţe şi respect faţă de cei cu care sociologul este pus în situaţia să colaboreze într-o cercetare.

2. Chestionarul

Am precizat de la început că instrumentul cu ajutorul căruia se realizează culegerea datelor în achete este chestionarul şi, prin urmare, va fi nevoie să ne oprim puţin asupra lui, spre a pune în evidenţă unele principii care stau la baza întocmirii sale, formele pe care le poate lua în diferite tipuri de anchetă.

Definiţia simplă de la care am plecat - chestionarul ca o succesiune de întrebări - are nevoie de cîteva adăugiri şi comentarii, deoarece este evident pentru oricine că nu orice set de întrebări poate primi numele de chestionar (sociologic) Dar nu doresc nici să caut o definiţie propriu-zisă, care să reziste unei analize serioase, căci mă tem că nu este posibil aşa ceva. Ceea ce mi se pare totuşi posibil şi necesar este faptul de a distinge chestionarul sociologic de alte forme asemănătoare, cu mare cîmp de circulaţie în viaţa de fiecare zi. Am spus chestionarul sociologic şi prin asta am introdus deja un element de delimitare, rezultat din însăşi finalitatea pentru care a fost construit instrumentul respectiv. Am senzaţia că oricît ne-am strădui, fără această notă, din care să rezulte scopul în care va fi folosit, este pînă la urmă imposibil să distingem un chestionar sociologic de o serie de documente prin care se transmite informaţia, de exemplu, între diferite niveluri ierarhice ale unei organizaţii sau de alte instrumente asemănătoare.

Prin urmare, chestionarul sociologic va cuprinde o serie de întrebări ce se adresează oamenilor care formează populaţia supusă investigării, întrebări ce joacă rolul de indicatori ai unor fapte sau fenomene sociale. Răspunsul primit la aceste întrebări - înregistrat de regulă în scris - furnizează o informaţie asupra cîmpului problematic cercetat, informaţie al cărei statut depinde de natura întrebărilor puse. Pentru precizarea acestui aspect este normal să reluăm clasificările tradiţionale ce se folosesc pentru întrebările chestionarelor.

I Mai întîi, după conținut , întrebările se pot împărți în trei mari categorii: factuale, de opinie și de cunoștințe.

Ia) Așa cum o arată și numele, întrebările factuale vizează obținerea unei informații legate de anumite evenimente petrecute în viața persoanei, a familiei sale, de unele însușiri obiective ce o caracterizează etc., deci de fapte, de stări care există așa cum există, independent de interpretările, opiniile pe care le poate avea omul în legătură cu ele. Numărul de copii, anul nașterii, retribuția din ultima lună, ce emisiuni a urmărit ieri la televizor, dacă are apă curentă în casă etc. - iată doar cîteva exemple de informații ce se urmăresc a se obține prin întrebările factuale. O caracteristică a acestor întrebări factuale o constituie posibilitatea - cel puțin în principiu - a verificării exactității răspunsurilor primite prin alte mijloace, cum sînt documentele, observația directă, chestionarea altor persoane etc.

Ib) Prin întrebările de opinie - și aici termenul de opinie trebuie luat într-un sens foarte larg - se urmărește accesul la aspectele de ordin subiectiv ale persoanei: atitudini, păreri, motive, interese etc. Tot acest univers psihologic al individului prezintă un interes deosebit pentru sociolog (și nu numai pentru cercetătorii din domeniul psihologiei), deoarece aspectele de conștiință nu rămîn izolate în sfera individuală, ci ele au un rol foarte important asupra comportamentului și acțiunilor oamenilor, fiind deci strîns legate de faptele propriu-zise. Mai mult, anumite concepții, păreri, atitudini etc., în cadrul social, se desprind de purtătorii individuali, deși ele nu există decît prin aceștia, devenind fapte sociale de sine stătătoare. Prin urmare, anchetele îndreptate în această direcție nu trebuie privite numai - sau în primul rînd - ca încercări de a descrie stările psihice individuale, ci unele elemente ale conștiinței colective, ca probleme detașabile, într-un anume sens, de indivizii de la care pornesc și prin care subzistă. Opiniile și atitudinile oamenilor în legătură cu munca sau cu alte aspecte ale vieții se concretizează în valori și comportamente sociale, legate de aceste elemente, valori și comportamente care la rîndul lor acționează asupra opiniilor și atitudinilor individuale.

Așadar, ne aflăm în fața unui univers problematic deosebit de complex, care ridică în fața cercetătorului multe obstacole

nu numai în abordarea dar şi în interpretarea fenomenelor studiate. Cît priveşte obstacolele metodice, ele se impun cu evidenţă oricărui cercetător şi se relevă mai simplu dacă vom compara situaţia întrebărilor de opinie cu cea a întrebărilor factuale. Mai întîi, despre anumite fapte ştim dacă este sau nu oportun să punem oamenilor întrebări, adică ştim dacă metoda anchetei poate duce la o informaţie consistentă. Il voi putea întreba liniştit pe fiecare om în legătură cu nivelul şcolar pe care l-a dobîndit, plecînd de la premisa normală că fiecare ştie cîte clase a urmat. In general universul întrebărilor factuale poate fi lărgit la toate aspectele despre care sîntem convinşi că individul uman deţine informaţia de care avem nevoie. Această delimitare este relativ facilă, singura problemă ce poate interveni fiind aceea dacă omul vrea să ne răspundă corect, adică se cer evitate acele întrebări pentru care există bănuiala că oamenii, dintr-un motiv sau altul (mai cu seamă de teama unor repercusiuni negative asupra vieţii lor). nu vor răspunde.

In schimb, în legătură cu atitudinile, opiniile etc. ar trebui aproape de fiecare dată să ne punem în mod serios problema dacă ele există, dacă are sau nu sens să adresăm oamenilor asemenea întrebări. De multe ori nu se ţine cont de acest lucru, considerîndu-se de la sine înţeles că putem chestiona o persoană despre, să zicem, satisfacţia pe care o are în muncă, părerea sau atitudinea faţă de un anumit eveniment social important şi astfel o parte însemnată din răspunsuri nu vor avea nici o acoperire, ele fiind date din complezenţă sau din teama omului că ar fi ridicol dacă ar recunoaşte cinstit că nu are nici o părere în legătură cu ceea ce-l întrebăm.

Un al doilea gen de dificultăţi metodologice legate de întrebările de opinie derivă din caracterul extrem de imprecis al problemei astfel urmărite, ceea ce face ca orice schimbare nu numai în textul întrebării dar chiar şi în tonalitatea cu care ea este adresată să producă modificări de accente şi deci de răspunsuri. Este motivul pentru care aproape în toate anchetele uzînd de astfel de întrebări se fac precizări exacte asupra modului cum se adresează întrebările de opinie şi se interzice operatorului orice schimbare, orice interpretare personală a formulării dată în

chestionar. Dacă la întrebările factuale operatorul are posibilitatea să uzeze, de cele mai multe ori, de orice intervenţie ce-o crede potrivită pentru a stimula subiectul să răspundă sau să găsească răspunsul corect, aici situaţia este complet diferită . Tot întrebările de opinie sînt cele care determină recurgerea la anchete directe pentru a nu da astfel timp de gîndire subiectului şi posibilitatea de a inventa sau căuta în jur opinii ce nu-i aparţin.

In sfîrşit, a treia problemă metodologică de bază legată de întrebările de opinie este ridicată de încercarea cercetătorului de a măsura intensitatea anumitor opinii, în speţă a atitudinilor. Psihologia socială a pus la punct un întreg arsenal în acest scop, dar spaţiul şi profilul cursului de faţă nu-mi permit să intru în detaliile acestei probleme complexe.

Inainte de a încheia discuţia despre întrebările de opinie aş dori să mai fac o precizare importantă, şi anume că opiniile oamenilor nu trebuie niciodată luate drept altceva decît opinii. Nu de puţine ori sociologul încearcă explicaţii sau soluţii la diverse probleme sociale şi atunci efectuează o anchetă de opinie spre a vedea care sînt părerile oamenilor legate de aspectele respective, avînd impresia că astfel va ajunge la ceea ce urmăreşte el. Nu este exclus ca din analiza acestor opinii să poată găsi explicaţia sau soluţia, dar, oricare ar fi situaţia, baza explicativă sau de rezolvare a problemei nu se găseste în opiniile exprimate. Chiar dacă oamenii cred în proporţie de 100% că un fenomen social negativ, să zicem, se datorează unei cauze anumite, sociologul nu are dreptul să se mărginească la o astfel de explicaţie. Aici, ca şi în multe alte probleme asemănătoare, ceea ce află el nu este decît interpretarea pe care o dau oamenii lucrurilor, interpretare ce poate fi corectă sau falsă şi care ea însăşi se cere interpretată şi explicată. Sociologul îşi pierde, evident, statutul de om de ştiinţă - la care ştim că aspiră atît de puternic - dacă transferă indivizilor din colectivitatea studiată responsabilitatea găsirii soluţiei problemelor sociale, iar sociologia s-ar transforma în acest caz într-o disciplină care ar culege explicaţii în loc să le furnizeze ea însăşi. Am dat exemplul explicaţiei şi cel al rezolvării problemelor sociale pentru că sînt cele mai elocvente în a susţine ideea de mai sus; dar ele nu sînt singurele chestiuni în

care opiniile trebuie tratate doar ca opinii, problema fiind, în același timp, generală dar și specifică, prin nunațe, fiecărui aspect studiat.

1c) Al treilea gen de întrebări, în funcție de conținutul lor, îl reprezintă cele de cunoștine. In acest caz, informația obținută de sociolog nu este legată de conținutul întrebării, ci de faptul dacă omul deține sau nu anumite cunoștințe. Dacă-l întreb la ce oră s-a sculat în dimineața asta, dorect să aflu efectiv acest lucru, tot așa cum dorect să-i evaluez, de pildă, gradul de satisfacție cu locuința în care trăiește. In schimb, dacă-l întreb cine a scris romanul "Ion" este clar că nu mă va interesa informația ca atare, ci coincidența răspunsului cu unul standard, considerat corect - "Rebreanu", în cazul de față - sau noncoincidența.

Așa cum bine se înțelege, rolul întrebărilor de cunoștințe este, de regulă, acela de a evalua, prin cîțiva indicatori foarte expresivi, nivelul de cultură (de fapt, al unor dimensiuni ale culturii) specific diferitelor grupuri umane. Adiacent, pot apare de aici și informații despre sistemul de valori, atitudini etc. , dacă interpretăm prezența sau absența cunoștințelor în anumite domenii. Problemele metodologice cele mai importante ridicate de întrevările de cunoștințe se referă la alegerea unor indicatori capabili să reflecte realmente nivelul în zona culturală investigată și apoi, mai pragmatic, de a găsi modalitatea cea mai potrivită de a adresa asemenea întrebări. Este evident că plăcerea omului de a răspunde la un chestionar, despre care am pomenit, dispare aproape complet atunci cînd el se simte "verificat" prin întrebări de cunoștințe. Din această cauză, se recomandă mqltă, foarte multă prudență în utilizarea acestor întrebări.

In rîndurile de mai sus am preluat o clasificare clasică a întrebărilor după conținut. Se poate observa că în unele manuale se vorbește despre o clasificare a chestionarelor după același criteriu. Consider însă că prea rare sînt situațiile cînd avem de a face cu chestionare "pure", formate numai din întrebări factuale, numai de opinie sau numai de conținut. De regulă, chestionarul cuprins de mai multe genuri de întrebări; în speță sînt aproape inexistente chestionarele formate doar din întrebări de opinie, acestora alăturîndu-li-se cel puțin cîteva întrebări factuale, cu

rol de clasificare a indivizilor după vîrstă, sex, nivel şcolar etc. Aceeaşi observaţie este valabilă şi pentru chestionarele axate precumpănitor pe întrebări de cunoştinţe. Chestionarele pur factuale pot avea ceva mai multe şanse de apariţie, dar şi aici, de obicei, sociologul este tentat să strecoare cîteva întrebări de opinie care să întregească universul informaţional cu elemente de apreciere, mulţumire etc. sau pur şi simplu pentru a introduce unele aspecte capabile să rupă monotonia unei anchete bazată numai pe date factuale.

A doua clasificare clasică a întrebărilor este cea privind

II forma de înregistrare a răspunsurilor. Se vorbeşte, în acest caz, de întrebări deschise, închise şi mixte (sau semiînchise ori semideschise).

IIa) Intrebările deschise sînt cele urmate în chestionar de un spaţiu liber în care se va înregistra răspunsul. In ancheta indirectă, subiectul fiind cel care completează chestionarul, va umple aşa cum crede el de cuvinţă acest spaţiu, adică va putea realiza fie un comentariu larg al problemei fie un rezumat scurt de cîteva cuvinte sau chiar de unul singur. Cînd se lucrează cu operatori de anchetă, aceştia sînt, de regulă, instruiţi să noteze răspunsul subiectului în mod cît mai detaliat şi mai fidel, interzicîndu-se deci rezumarea sau interpretarea personală din partea operatorului. Desigur că o asemenea cerinţă rămîne de cele mai multe ori un deziderat greu de împlinit şi chiat nerealizabil, avînd în vedere multitudinea situaţiilor întîlnite pe teren. In orice caz, se pare că cel puţin o rezumare a răspunsurilor ample este inevitabilă, operatorul trebuind să se străduiască să nu piardă nici una din ideile principale şi să nu le altereze sensul.

IIb) Intrebările închise sînt cele însoţite de variante posibile de răspuns, subiectul sau operatorul fiind obligaţi să aleagă pe cea potrivită. Spun subiectul sau operatorul deoarece se pot ivi ambele situaţii. Subiectul este chemat să aleagă varianta de răspuns întotdeauna în ancheta indirectă şi uneori în cea directă, atunci cînd după citirea întrebării operatorul are sarcina să anunţe şi variantele şi să roage subiectul să aleagă varianta în conformitate cu opinia sau faptul în cauză. Alteori în ancheta directă subiectului nu i se aduc la cunoştinţă variantele de răspuns, între-

barea avînd pentru acesta formă deschisă, dar operatorul încadrează răspunsul în una din variantele prevăzute.

Iată cîteva exemple de întrebări închise:

A. Care este vîrsta dumneavoastră ?

1. Sub 2o ani
2. 2o - 29 ani
3. 3o - 39 ani
4. 4o - 49 ani
5. 5o - 59 ani
6. 6o - 69 ani
7. 7o ani şi peste

B. In ce măsură sînteţi satisfăcuţi de numărul de camere din locuinţa dv. ?

1. Foarte mult
2. Mult
3. Moderat
4. Puţin
5. Foarte puţin

C. Care sînt modalităţile preferate de dv. pentru petrecerea timpului liber ?

1. Lectură
2. Cinematograf
3. Televizor
4. Excursii, plimbări
5. Practicarea sportului
6. Asistenţă la manifestări sportive
7. Dans
8. Teatru
9. Operă

Răspunsul la întrebările închise se realizează prin marcarea variantei de răspuns aleasă în diferite forme: fie încercuind cifra de cod care, de regulă, precede varianta respectivă, fie trecînd aceeaşi cifră într-o căsuţă (pătrăţel) construit alături de întrebare, pe marginea foii, spre a ieşi mai clar în evidenţă, fie prin alte metode (subliniere, marcare cu X etc.) Variantele de răspuns pot fi sau nu exclusive. Astfel, în primele două exemple de întrebări este clar că nu se poate marca decît o variantă, pe cînd în a treia se pot alege două, trei sau chiar mai multe.

Una dintre problemele tehnice cele mai importante legate

de construcţia întrebărilor închise este cea care priveşte epuizarea întregului univers referenţial prin variantele alese, adică cercetătorul trebuie să prevadă, din etapa proiectării investigaţiei, toate posibilităţile de răspuns. Din cele trei exemple alese, primul nu ridică nici o problemă de acest gen. Al doilea, aparent şi el surprinde, prin forma de scală ordinală, un evantai satisfăcător de variante; o singură chestiune poate apărea în cazuri ca acesta, şi anume imposibilitatea subiectului de a emite opinia cerută. De aceea, pentru a nu forţa subiectul să dea un răspuns neconform cu realitatea, este de multe ori preferabil că acestor variante de răspuns să li se mai adauge una din genul: "nu ştiu", "nu mă pot pronunţa", "nu m-am gîndit la problema aceasta" etc.

Intrebarea C am pus-o intenţionat într-o formă defectuoasă din punctul de vedere al surprinderii tuturor posibilităţilor de răspuns. Este evident că lista propusă acolo trebuia continuată cu încă alte variante de petrecere a timpului liber, dar tot atît de adevărat este că la întrebarea noastră pot apare cele mai neaşteptate răspunsuri şi, prin urmare, e greu de a respecta cerinţa exhaustivităţii chiar dacă lista s-ar îmbogăţi substanţial In aceste condiţii, se recurge la cea de a treia formă de întrebări -
II c) mixte sau semiînchise (semideschise) - ce cuprind, pe lîngă setul de variante explicitate, încă una din genul "Alte situaţii. Care? ... " Se cere astfel atît încercuirea unui cod - ataşat acestei variante - dar şi exprimarea liberă a situaţiei neprevăzută în variantele scrise.

Cît priveşte frecvenţa de utilizare a celor trei tipuri de forme de înregistrare a răspunsurilor, se poate emite următoarea apreciere. Intrebările factuale sînt aproape în totalitate redactate sub formă închisă, întrucît variantele de răspuns sînt mai uşor previzibile. Doar atunci cînd numărul alternativelor este prea mare, după înşiruirea cîtorva, socotite principale (la care se aşteaptă frecvenţa de răspuns mai ridicate) se poate apela la varianta "Alte situaţii". De pildă, la întrebările obişnuite despre ocupaţie, putem folosi următoarea schemă:

1. agricultor
2. muncitor
3. maistru
4. tehnician, proiectant, subinginer

5. funcţionar
6. lucrător în servicii
7. intelectual
8. alta . . . .

In cazul întrebărilor de opinie, distingem în special clasa acelora în care se urmăreşte evaluarea intensităţii unei atitudini, aprecieri, ataşament la o valoare etc., caz în care se utilizează pentru răspuns o scală ordinală închisă, ca la întrebarea B. Treptele scalei pot fi, cum sînt acolo, în număr de 5, mai mic sau mai mare.

Iată încă două exemple:

D. Cum apreciaţi emisiunea X, din programul televiziunii din ziua. ... ?

1. Bună
2. Mediocră
3. Slabă

E. In ce măsură sînteţi mulţumiţi de modul de gospodărire a localităţii în care trăiţi ?

1. Foarte mult
2. Mult
3. Relativ mult
4. Nici mult, nici puţin
5. Relativ puţin
6. Puţin
7. Foarte puţin

Evident, ordinea variantelor poate fi cea dată aici sau una inversă, începînd deci cu aprecierile negative. Această din urmă situaţie se întîlneşte mai ales atunci cînd numărul din faţa variantei este luat nu pur şi simplu ca simbol de cod ,ci drept măsură a intensităţii opiniei respective. La limită, în asemenea cazuri se poate ajunge să nu se mai treacă deloc variantele de răspuns pe chestionar, ci punînd de exemplu întrebarea E, să se precizeze: "daţi o notă de la 1 la lo, care să reflecte opinia dv., convenind că 1 corespunde insatisfacţiei totale iar lo satisfacţiei maxime".

Restul tipurilor de întrebări de opinie - care nu se pretea-

ză la o astfel de scalare a răspunsurilor -, exceptînd cele care cer răspunsuri în cadrul dihotomiei "Da/Nu" şi care, la limită, pot fi considerate ca aparţinînd tipului de mai sus, restul întrebărilor deci pot fi mai greu surprinse sub forma evantaiului închis de răspunsuri şi ele vor apare, în bună măsură, sub formă deschisă sau, eventual, semiînchisă.

In ceea ce priveşte întrebările de cunoştinţe, ele se pot prezenta în egală măsură în cele două tipuri formale principale. Deşi aparent forma deschisă pare cea mai potrivită pentru ele, adesea se foloseşte un evantai de răspunsuri prestabilite, cerîndu-i-se subiectului să aleagă varianta corectă (aşa cum se procedează şi în chestionarele folosite la examenul pentru obţinerea carnetului de conducere auto). Această modalitate din urmă este utilizată cu deosebire în ancheta indirectă (autocompletare) pentru a evita amestecul deprinderilor de scris şi cunoştinţelor de limbă cu cele urmărite de cercetare. Un exemplu de astfel de întrebare poate fi:

F. Cine este autorul romanului "Ion" ?

1. Sadoveanu
2. Agîrbiceanu
3. Rebreanu

Recapitulînd şi încercînd o evaluare globală, putem constata că întrebările închise sînt cele care predomină de departe în marea majoritate a chestionarelor sociologice. Raţiunile pentru care se întîmplă acest lucru ar putea fi sistematizate astfel:

Mai întîi - şi cu siguranţă, că aceasta este motivul fundamental - la egalitatea celorlalţi factori, întrebările închise sînt preferate pentru facilitatea prelucrării informaţiei obţinute. Orice formă ar lua răspunsurile la o întrebare, pînă la urmă ele trebuie reduse şi aduse la nişte elemente standardizate care să ofere posibilitatea comparării informaţiei obţinute de la diferiţi indivizi, ceea ce înseamnă că şi întrebările libere trebuie , în ultimă instanţă "închise" într-un sistem clasificatoriu. Or, asemenea muncă este extrem de migăloasă şi, aşa cum au arătat unele experimente metodologice, ea este acompaniată de un procent ridicat de erori de codificare. In esenţă, prelucrarea întrebărilor deschise este o veritabilă analiză de conţinut şi se înţelege că

dacă numărul acestor întrebări este mare, efortul solicitat de sistematizarea informaţiei e atît de ridicat încît poate depăşi cu uşurinţă capacităţile echipei de cercetători. La întrebările închise, rubricile de prelucrare sînt gata elaborate astfel că adesea informaţia se trece direct de pe chestionar în sistemul automat de prelucrare a datelor. Intrebările semiînchise pot ridica unele probleme la prelucrare dacă frecvenţa răspunsurilor la varianta liberă (nedeterminată) este mare, ceea ce înseamnă că s-au omis unele aspecte ce nu pot fi neglijate şi este nevoie deci ca ele să fie puse în evidenţă prin analiza răspunsurilor concrete.

Al doilea motiv ce poate fi invocat în favoarea întrebărilor închise este legat de facilitatea mînuirii unui chestionar în care predomină întrebările de acest gen, deci de posibilităţile rapide de completare, fapt ce-i asigură o eficienţă practică deosebită.

In al treilea rînd, întrebările închise prezintă şi un avantaj legat de ajutorul pe care variantele de răspuns îl dau subiectului în înţelegerea corectă a sensului întrebării din chestionar. Acest lucru este valabil mai ales în anchetele indirecte, cînd anumite defecţiuni în formularea întrebării sau ambiguitatea unor termeni pot conduce la interpretarea neuniformă a unei întrebări. Or, văzînd care sînt posibilităţile de răspuns, subiectul este într-un fel dirijat pe făgaşul urmărit de anchetator.

In sfîrşit, o problemă apropiată de precedenta, dar nu identică, ar fi aceea că variantele de răspuns pot "provoca" o anumită opinie sau reactualiza o informaţie altfel ignorată. De multe ori omul nu are clar conştientizată o anumită atitudine sau opinie, pe care însă subiectul o regăseşte imediat ca fiind a lui atunci cînd citeşte sau i se citesc variantele unei întrebări. Ideea merge, cum spuneam, şi la întrebările factuale. De exemplu, dacă cerem omului data unui eveniment mai îndepărtat în timp el nu va fi capabil s-o spună cu exactitate; în schimb, găsind nişte variante de răspuns construite pe intervale de timp delimitate de anumite evenimente istorice, el nu va avea nici o dificultate să încadreze corect evenimentul.

Cu acest aspect al problemei ajungem şi la dezavantajele

întrebărilor închise. Aşa cum mai spuneam, omul nu are întotdeauna opinii formate şi o întrebare închisă îl poate ajuta să iasă cu faţa curată, indicînd la întîmplare una dintre variante. Alte persoane, dimpotrivă, se vor simţi stînjenite de unele delimitări categorice aşa cum apar prin variantele indicate, considerînd că un răspuns liber ar fi mai adecvat , le-ar permite să-şi exprime mai nuanţat părerile. De asemenea, un chestionar bazat aproape în exclusivitate pe întrebări închise poate crea o atmosferă de "răceală" între operator şi subiect, de rigiditate sau chiar de superficialitate în răspunsuri.

Unele dintre neajunsurile întrebărilor închise sînt eliminate în ancheta directă, atunci cînd operatorul nu aduce la cunoştinţă subiectului variantele de răspuns, ci el este cel care o alege pe cea potrivită în urma unui răspuns liber. In orice caz, întrebările închise s-au statornicit cu multă forţă în chestionarele sociologice şi este evident că acest lucru s-a întîmplat tocmai datorită avantajelor lor incontestabile. Intrebările libere punctează din loc în loc chestionarul pentru a face discuţia mai atractivă, pentru a stimula subiectul cu anumite probleme deosebite, şi desigur, pentru a căuta nuanţe şi informaţii suplimentare la acele chestiuni importante ce nu pot fi expediate în cîteva sumare variante de răspuns.

III O altă clasificare a întrebărilor, ce se cuvine amintită aici, este cea care distinge întrebările directe de cele indirecte. In adevăr, dat fiind faptul că omul nu este întotdeauna dispus, din motive extrem de diverse, să dezvăluie o anumită informaţie pe care cercetătorul o aşteaptă, acesta din urmă este pus în situaţia de a face cîteodată apel la unele căi ocolitoare de atingere a scopului urmărit. O întrebare va fi numită indirectă dacă sensul real al său sau al răspunsului este diferit de cel aparent, adică dacă prin intermediul său se urmăreşte un lucru care nu este sesizat de cel interogat. Am amintit mai înainte că prin unele întrebări de cunoştinţe putem urmări nu atît ce şi cît cunoaşte un om, ci profilul unui gen de cunoştinţe spre a ne dezvălui eventuale orientări valorice. Există multe genuri de probleme sociologice care pot fi studiate doar prin astfel de întrebări indirecte, cum ar fi, de pildă, gradul de religiozitate, starea materială, comportamentul sexual etc. Este limpede că nu vom adresa

oamenilor direct întrebări de tipul: "Credeţi în Dumnezeu?", "Cîţi bani cîştigaţi pe lună în afara retribuţiei ?" sau altele la care răspunsurile vor fi evident distorsionate de teama anumitor consecinţe negative, de tendinţa indivizilor de a se conforma - formal - unor norme şi valori sociale dominante, în general de sistemul de autoapărare pe care fiecare şi-l formează în relaţiile cu semenii săi.

La încheierea acestui paragraf, doresc să mai spun doar cîteva cuvinte despre numărul şi ordinea întrebărilor din chestionar. Mai întîi o precizare: în ciuda impresiei care poate persista în rîndul unor nespecialişti, cum că atîtea decenii de cercetare sociologică empirică ar fi condus la găsirea unor soluţii optimale, general valabile, în legătură cu cele două caracteristici ale chestionarului menţionate, nu există totuşi o reţetă cu aplicabilitate universală nici în privinţa numărului şi nici a ordinii întrebărilor.

Chestionarele pot conţine de la una-două întrebări la 3oo sau chiar mai multe şi e imposibil de spus dinainte ce număr este preferabil. Acesta depinde de o mulţime de factori care ţin de conţinutul problemei de studiat, de complexitatea faptelor şi fenomenelor urmărite, de bogăţia interacţiunilor cu altele, mai ţin de cunoştinţele pe care cercetătorul le are despre aceste fenomene, de consistenţa schemelor teoretice şi explicative de care dispune, dar şi nu în ultimul rînd - de aspecte exterioare, în speţă de mijloacele materiale şi timpul avut la dispoziţie. In plus, numărul întrebărilor va depinde de tehnica de aplicare a chestionarului (o anchetă prin poştă va face apel la chestionar mai scurt decît una directă) şi de populaţia căreia i se adresează (în general, persoanele cu studii superioare sînt mai uşor de antrenat în discuţii pe baza unor chestionare lungi sau le pot completa şi singure cu mai multă uşurinţă).

Referitor la ordinea întrebărilor, se poate aprecia că aceasta capătă o importanţă deosebită în următoarele două cazuri:

a) cînd conţinutul unor întrebări este de aşa natură încît ele trebuie aşezate la distanţă cît mai mare una de alta pentru ca răspunsurile să nu se influenţeue reciproc şi

b) cînd chestionarul conţine unele întrebări "dificile" care pot trezi neîncrederea subiectului sau genera refuzul de a mai răspunde; astfel de întrebări se vor aşeza, evident, spre sfîrşitul chestionarului pentru a diminua pe cît posibil efectele negative.

Tot în finalul chestionarelor în care predomină întrebările de opinie se recomandă — mai ales dacă subiectului i se garantează anonimatul — plasarea întrebărilor factuale de clasificare, întrebări care uneori pot lăsa impresia că prin ele se urmăreşte culegerea cît mai multor indici pentru identificarea ulterioară a persoanei care răspunde. Este bine ca, ajunşi în acest punct cu interviul, operatorii să lămurească exact rostul întrebărilor de clasificare în chestionar.

3. Erori şi surse de erori în anchete

Datorită faptului că ancheta este o metodă de investigare în cursul desfăşurării căreia intervin nenumăraţi factori perturbatori, manualele de metode şi tehnici se străduiesc, în mod cu totul îndreptăţit, să invantarieze cît mai complet sursele care provoacă erori, în diferitele faze ale cercetării. O idee de principiu ce se cere subliniată de la început este aceea că, de regulă, o eroare nu este generată de un singur factor, ci apare ca rezultat al interacţiunii tuturor factorilor ce intervin la momentul respectiv al cercetării. Cu toate acestea, obiceiul este să se treacă în revistă erorile după factorul principal generator, ceea ce voi face şi eu în continuare, dar ideea de mai sus trebuie să rămînă mereu prezentă în minte spre a evita riscul unor interpretări unilaterale.

3.1. Erori generate de construcţia chestionarului

Deja din cele spuse pînă aici s-au degajat cîteva aspecte de construcţie a chestionarului generatoare de erori, cum ar fi conţinutul întrebărilor, lungimea excesivă a chestionarului sau ordinea nepotrivită a întrebărilor, aspecte asupra cărora nu voi mai reveni în continuare.

O problemă, implicînd de fapt ordinea întrebărilor, dar cu o semnificaţie mult mai largă şi cu un rol foarte important în producerea erorilor, mai cu seamă dacă e vorba de anchete indirecte, este structura (logică şi grafică) a chestionarului. Mă gîndesc aici în special la faptul că un chestionar este rareori constituit din întrebări care toate se adresează întregului lot anchetat. Adică, altfel spus, apar frecvent întrebări ce nu se vor adresa anumitor persoane pentru că ele nu vizează situaţii specifice acestora. De exemplu, dacă studiem calitatea vieţii într-o colectivitate, vom ajunge inevitabil să abordăm şi aspecte legate de muncă, de locul de muncă, de transportul spre acesta etc. Or, este evident că o parte a populaţiei, fiind inactivă, nu va fi vizată de asemenea întrebări. In toate aceste cazuri, chestionarul trebuie, prin menţiuni foarte clare şi vizibile, să pună în evidenţă ce categorie de persoane răspunde la o anumită întrebare. Apar astfel în chestionare aşa-numitele întrebări filtru care orientează subiectul spre întrebările la care să răspundă. In exemplul de mai sus, o asemenea întrebare filtru, avînd, să zicem numărul 30, va fi următoarea:

30. Sînteţi încadrat în cîmpul muncii ?
 1. Da
 2. Nu

Presupunînd că după ea urmează 10 întrebări referitoare la persoanele active (numerotate de la 31 la 40), va fi necesar ca întrebarea filtru, cu nr.30, să fie însoţită de menţiunea: "Cei care răspund 'da', vor continua cu întrebarea 31, iar cei care răspund 'nu', vor trece direct la întrebarea 41". Sau - şi mai bine - întrebările 31-40 vor fi puse într-un chenar deasupra sau în dreptul căruia se va specifica în mod foarte vizibil: "Numai pentru persoanele care au răspuns 'Da' la întrebarea 30, deci cele încadrate în cîmpul muncii".

Desigur, sînt şi alte modalităţi de vizualizare a acestor bifurcaţii şi nu este niciodată de prisos să se ia cît mai multe măsuri de precauţie, chiar şi atunci cînd se lucrează cu operatori de interviu, instruiţi în prealabil. Erorile de această natură sînt practic prezente în orice anchetă indirectă căci este aproape imposibil ca oamenii să nu completeze şi întrebări care nu-i pri-

la interpretări eronate ale răspunsurilor subiecţilor, care vor fi scrise ca atare, la întrebările libere sau traduse în coduri, la cele închise. O sursă interesantă de erori, pusă în evidenţă de unele studii metodologice - vezi Moser (1967, pp. 356-357) - o reprezintă aşa numitele anticipaţii ale operatorilor de anchetă.

Operatorul poate fi influenţat în completarea chestionarului de răspunderile pe care le primeşte la primele întrebări,de unde el îşi face o imagine asupra atitudinii de ansamblu a subiectului în legătură cu problema studiată, avînd deci tendinţa să interpreteze orice răspuns în lumina acestei imagini-model (anticipaţiile de structură-atitudine). Pe de altă parte, luînd contact cu persoana de anchetat, operatorul observă unele elemente ce ţin de vîrstă, condiţia socială şi intelectuală a ei, de starea materială etc. şi deci situează această persoană pe o anumită treaptă socială, o identifică cu un anumit tip social. In consecinţă, orice ambiguitate de răspuns, orice neclaritate va fi tranşată în sensul imaginii respective, adică în direcţia unor răspunsuri aşteptate de la acest personaj social (anticipaţiile de rol). In fine, dacă înaintea anchetei operatorul are o anumită ipoteză asupra distribuţiei răspunsurilor la unele întrebări (de exemplu va considera că trebuie să găsească un procent mare de nemulţumiţi în legătură cu o problemă oarecare), atunci el va încerca, probabil, ca toate răspunsurile ambigue să le claseze de aşa manieră încît să se încadreze în imaginea lui iniţială. Un alt aspect, consider eu de aceeaşi natură, întîlnit frecvent în anchetele realizate cu studenţi în calitate de operatori,apare cînd, după realizarea unui număr de chestionare, se constată că la anumite întrebări răspunsurile se dovedesc a fi uniforme, adică o variantă este aleasă de un procent foarte ridicat de indivizi. In asemenea situaţii, operatorul are tendinţa în continuare să claseze identic răspunsurile celorlalţi subiecţi. In ambele cazuri se poate vorbi de anticipaţie de probabilitate. Tot din experienţa personală cu operatorii studenţi pot mărturisi că uneori aceste anticipaţii, mai ales forma ultimă menţionată, sînt atît de puternice încît operatorul nici nu se mai străduieşte să adreseze unele întrebări din chestionar, completînd automat răspunsul. Probabil că în 99% din cazuri el nu greşeşte, dar niciodată nu sîntem siguri ce va spune

un om şi deci o asemenea practică trebuie considerată, în principiu, ca generatoare de erori.

In legătură cu modalitatea de adresare a întrebărilor este greu de a detecta şi clasifica riguros toate aspectele ce pot provoca erori. Uneori, la întrebările de opinie, este dificil de precizat dacă un anume fel de a te adresa conduce la eroare sau mai degrabă la un răspuns corect. Aşa cum am mai spus, răspunsul la aceste întrebări este foarte sensibil la modul de formulare şi d de aceea în multe anchete se cere operatorilor utilizarea unei forme standard, spre a se reduce măcar variabilitatea răspunsurilor datorată acestui fapt. Dar o astfel de formă nu este şi nu poate fi niciodată respectată riguros de toţi operatorii, esenţialul rămînînd ca abaterile individuale să se compenseze reciproc.

3.3. Erori datorate subiectului şi contextului social în care se realizează ancheta

Probabil că dintre toate erorile ce afectează o cercetare, cele mai puţine sînt cele ce pot fi puse precumpănitor în seama anchetatorului, căci, aşa cum judicios remarcă Moser: "ca şi în viaţa de toate zilele, ceea ce oamenii spun şi cum spun depinde de împrejurări şi de persoana cu care vorbesc" (1967, p.358). Am putea totuşi , într-o primă instanţă, clasifica erorile datorate subiectului în două categorii: conştiente şi inconştiente, sau poate mai bine spus, intenţionate şi neintenţionate.

Fiecare din aceste tipuri se regăseşte atît în cazul întrebărilor factuale cît şi în cel al întrebărilor de opinie. Dacă omul nu doreşte să dezvăluie anumite stări de fapt (de exemplu, veniturile) sau anumite opinii el va da răspunsuri nu în conformitate cu situaţia reală, ci cu una ipotetică pe care o consideră a fi cea mai adecvată în contextul respectiv (va minimaliza veniturile, va declara opinii care să se încadreze în modelul dezirabilităţii sociale etc.) Erorile neintenţionate pot fi datorate înţelegerii greşite a conţinutului întrebării, faptului că nu-şi aminteşte corect anumite evenimente etc.

Erorile ce provin din răspunsurile neadecvate ale subiectului se datorează în bună măsură structurii psihice a acestuia. Memoria, atenţia, voinţa, curajul, lipsa de prejudecăţi etc. sînt

cîţiva din factorii de personalitate ce-şi pun amprenta asupra comportamentului subiectului în situaţia de interviu. Nu este de neglijat nici influenţa contextului social în care se realizează ancheta. Anumite presiuni sociale de moment pot produce distorsiuni serioase în răspunsuri. De altfel, problema aceasta este ceva mai generală; anchetele sociologice sînt ele însele fapte sociale cu care oamenii sînt mai mult sau mai puţin familiarizaţi. De regulă, într-o societate în care nu există tradiţia unor asemenea investigaţii, persoanele sînt suprinse şi bănuitoare atunci cînd se trezesc în faţa "omului cu chestionarul". Este nevoie deci de consolidarea unei experienţe sociale, de înţelegerea de către toată lumea că anchetele sociologice sînt inofensive faţă de un individ sau altul ca persoană privată, că ele slujesc în societatea noastră unor scopuri nobile, de cunoaştere şi de ameliorare a vieţii tuturor.

3.4. Erori introduse în faza de prelucrare a datelor

Două ar fi sursele de erori în faza finală a unei anchete sociologice. Mai întîi, erorile pot apare din greşelile săvîrşite în operaţia de postcodificare a întrebărilor deschise. Am amintit deja cît de multe dificultăţi se ridică la prelucrarea acestor întrebări; oricît de riguroasă ar fi mulţimea categoriilor stabilite pentru încadrarea răspunsurilor respective, este practic imposibil să se asigure prelucrări o fidelitate foarte ridicată. Aşa cum o demonstrează cu claritate rezultatele experimentelor efectuate în acest sens, repetarea codificării informaţiei conţinută în întrebările deschise, cu aceeaşi grilă de categorii dar de către alte persoane - sau chiar de o aceeaşi persoană după un timp oarecare - conduce la rezultate sensibil diferite (vezi Moser, 1967, pp. 388 - 393).

O altă operaţie efectuată în faza fianlă a cercetării, care este întotdeauna generatoare de erori, este cea de transpunere a informaţiei de pe chestionar în instrumentele de prelucrare a datelor. Astfel, dacă chestionarul se prelucrează la calculator, toate codurile sale pot fi trecute iniţial pe un formular standard după care se perforează cartelele; uneori, dacă chestionarul nu es-

te foarte complicat şi mai ales dacă sistemul de codificare e uniform şi clar, codurile pot fi trecute direct din chestionar pe cartele. Deşi riscul de a greşi, pentru persoanele ce perforează cartelele, este mai mare cînd lucrează direct cu chestionarul, eliminarea unei operaţii intermediare conduce cu siguranţă şi la eliminarea multor erori. De asemene , şi în cursul prelucrării manuale a datelor este posibil să se introducă erori; din fericire, aici ele pot fi descoperite relativ uşor şi corijate aşa precum, în cursul acestei prelucrări, mai pot fi detectate şi corectate unele greşeli introduse în fazele anterioare ale cercetării.

Se poate aprecia că în faza prelucării datelor - manuală sau automatizată -, expeptînd desigur cazurile de neglijenţă crasă, erorile sînt relativ puţine şi nu pot avea o influenţă semnificativă asupra rezultatelor. Ba, mai mult, dacă dorim o precizie foarte mare, cu eforturi nu tocmai deosebite erorile pot fi aproape complet eliminate. De pildă, la prelucrarea la calculator, cea mai simplă şi eficace măsură de a elimina greşelile intervenite în transpunerea informaţiei pe cartele constă în perforarea, direct de pe chestionar , a două rînduri de cartele - dacă e posibil de către operatoare diferite - şi compararea, tot cu ajutorul ordinatorului, a perechilor de cartele corespunzînd aceluiaşi chestionar; se va pune imediat în evidenţă orice deosebire între două cartele paralele, identificîndu-se astfel erorile.

x

x x

Este clar că o serie de erori ce intervin în faza de prelucrare a datelor îşi au izvorul în modul în care informaţia este înregistrată pe chestionar. Pentru a evita asemenea greşeli şi chiar pentru a corecta unele erori introduse în faza de culegere a informaţiei, înainte de prelucrare, chestionarele trebuie supuse unei operaţii riguroase de verificare. Sub nici un motiv, chiar dacă se dispune de anchetatori foarte pricepuţi, chestionarele nu vor fi trimise la prelucrare fără ca un grup, cît mai restrîns posibil, de cercetători să treacă la o verificare a lor. In cursul acestei

operaţii se urmăresc, în esenţă, trei lucruri de bază: completitudinea, exactitatea şi uniformitatea înregistrării datelor. (Pentru detalii, a se vedea Moser, 1967, pp.379-383). Se vor elimina astfel erori însemnate datorate unor omisiuni în transcrierea informaţiei, unor inadvertenţe în răspunsurile la întrebările ce se controlează reciproc şi a unor interpretări neuniforme a sensului anumitor întrebări. Majoritatea acestor erori, depistate înaintea prelucrării, pot fi eliminate pe cale logică sau prin reanchetarea unor subiecţi.

Trecînd în evidenţă sursele erorilor în anchetele sociologice, se poate crea impresia că datorită numărului foarte mare al acestora şi a intensităţii cu care unele acţionează ar trebui să punem serios sub semnul întrebării eficienţa metodei, validitatea rezultatelor obţinute. Desigur, erorile în orice anchetă au o pondere deloc neglijabilă şi nu e cazul nici să idealizăm metoda trecînd sub tăcere aspectele care nu ne plac. Totuşi, acesta nu e un motiv suficient de puternic pentru a abandona una din metodele de bază ale sociologiei. Mai întîi, pentru că, dacă o cercetare este condusă cu seriozitate, erorile nu pot fi atît de mari încît să denatureze total sensul şi semnificaţia unor constatări la care ea ajunge. Chiar dacă proporţiile sau mediile unor mărimi determinate prin anchetă nu sînt identice cu cele "reale" abaterile nu vor depăşi o limită de toleranţă acceptabilă. Găsind, de exemplu, că 75% dintre indivizii chestionaţi îşi exprimă insatisfacţia faţă de un anumit aspect al vieţii lor, înseamnă chiar că ne găsim în faţa unei probleme sociale ce reclamă o rezolvare rapidă, chit că procentul real nu e 75, ci poate 7o sau 8o. Importă mai puţin valoarea exactă a unui asemenea procent şi mai mult ordinul său de mărime asupra căruia nu se poate chiar atît de uşor greşi.

In al doilea rînd, realizarea anchetei pe o masă mare de indivizi face posibilă acţiunea legităţilor statistice, nu numai în scopul desprinderii unor tendinţe în populaţie, dincolo de comportamentele individuale, dar şi referitoare la erorile accidentale, nesistematice. Vreau să spun că, pe astfel de populaţii mari, o bună parte din erorile efective produse în fiecare chestionar, numite erori brute, se vor elimina prin anihilări reciproce, în calculul unor mărimi medii la nivelul lotului cercetat,

şi vor rămîne erori finale mai reduse, numite erori nete.

In concluzie, se poate deci aprecia că problema erorilor în cercetarea de tip anchetă este una fundamentală, căreia sociologul trebuie să-i acorde întreaga atenţie, de la începutul şi pînă la sfîrşitul cercetării; nici un efort în acest sens nu va rămîne nerăsplătit. Intr-o asemenea strategie generală locul central trebuie să-l ocupe depistarea factorilor ce produc erori sistematice, a căror eliminare nu este de aşteptat din partea hazardului; ei sînt legaţi, aşa cum s-a văzut, în special de modul de întocmire a chestionarului şi de munca operatorilor.

Cap.V. ANALIZA DOCUMENTELOR

O altă metodă specifică sociologiei - şi în general ştiinţelor socio-umane, - alături de anchetă, o constituie tot o specie derivată din metoda generală a observaţiei, aplicată însă nu la realitatea socială în sine, ci la o imagine a ei, şi anume la documentele produse şi reproduse în fiecare tip de societate umană. Vom numi metoda cu termenii din titlu: analiza documentelor.

Lecţia de faţă urmăreşte să prezinte trei probleme fundamentale: ce, unde şi cum se investighează cu metoda în cauză, adică să specifice conţinutul documentelor care constituie obiectul analizei sociologice, tipul de documente în care acest conţinut este identificabil şi tehnicile de analiză subordonate metodei. Primele două aspecte vor fi schiţate împreună, dat fiind că despre conţinutul documentelor cu greu s-ar putea vorbi în abstract, fără a face referiri la natura acestora, din cauza enormei varietăţi de informaţii documentare ce sînt utile sociologului, în nenumărate probleme studiate de el.

Cît priveşte rostul şi importanţa metodei în ştiinţele sociale, nu cred că este nevoie să se spună prea multe; oricine înţelege uşor că informaţia documentară este una absolut indispensabilă "observatorului" faptelor sociale, că neglijarea ei şi reducerea arsenalului metodologic al sociologiei doar la observaţii directe sau anchete înseamnă renunţarea cu bună ştiinţă la o serie de date şi informaţii ce nu pot fi reconstituite cu alte mijloace. De altfel, se poate aprecia că analiza documentelor a constituit prima metodă majoră de cercetare în sociologia nespeculativă. Lucrări clasice în domeniu precum Capitalul lui Marx, Situaţia clasei muncitoare din Anglia a lui Engels, Le Suicide de Durkheim sau scrierile lui Lenin despre dezvoltarea capitalismului în Rusia, ca să pomenesc doar cîteva exemple mai semnificative, sînt în întregime elaborate în substanţa lor informaţională cu ajutorul metodei şi prezentată.

Odată cu dezvoltarea sociologiei empirice de teren, alte

metode au trecut în prim-planul atenției sociologilor: ancheta, interviul sau observația. Cu toate acestea, se poate afirma că analiza documentară n-a fost niciodată abandonată chiar dacă a fost folosită numai ca însoțitoare a altor metode și asta cel puțin pentru un motiv esențial: înțelegerea fenomenelor prezente nu este, de regulă, satisfăcătoare fără a cunoaște evoluția ori geneza lor, deci trecutul, formele de manifestare pînă în momentul cercetării. Or, despre trecut, metoda care furnizează datele cele mai consistente este tocmai analiza documentară.

După părerea mea, la ora actuală, cuplul metodologic anchetă- analiza documentelor constituie instrumentul cel mai eficace de cunoaștere pentru sociologie. Desigur, sînt situații în care numai una dintre cele două metode poate fi utilizată, dar nu este mai puțin adevărat că deseori ancheta este pornită pe un teren gol, nesusținută de alte surse de cunoaștere, și datele furnizate de ea sînt inconsistente, ajungîndu-se astfel la discreditarea acestei importante metode de cunoaștere a vieții sociale.

Legitimitatea analizei documentelor, ca metodă sociologică, se află, în ultima instanță, în faptul că documentele - indiferent de natura lor - sînt produsul activității umane, produs în care se încorporează o anumită cantitate de informație despre faptele și fenomenele sociale din perioada în care ele au fost produse. La prima vedere poate părea că metoda ar fi proprie doar istoriei și nu sociologiei căci documentele dau seama totdeauna despre un trecut. Argumentele împotriva acestei idei sînt însă imediate:

(i) Sociologia, în accepția multora dintre cei care o practică, nu este numai știința prezentului social, ci și a trecutului, fiind complementară studiului istoric al societăților pe axa cronologică.

(ii) In al doilea rînd, așa cum aminteam deja, cunoașterea trecutului este deseori indispensabilă pentru înțelegerea corectă a prezentului.

(iii) Chiar dacă nu acceptăm ideile de mai sus în toată amploarea lor, va trebui totuși admis că prezentul nu se reduce la o clipă ce trece rapid, ci are și el o durată, deci un trecut mai lung sau mai scurt ce-i aparține în fapt. Sau, altfel spus, docu-

mentele nu sînt caracteristice numai trecutului, ci şi prezentului; ele dau seama, uneori foarte bine, despre ceea ce se întîmplă acum şi aici.

(iv) In fine, metoda documentară nu ţine doar de studiul istoric chiar şi pentru simplul motiv că ea se aplică, aşa cum vom vedea imediat, chiar la "obiecte" create de cercetarea sociologică (documente constituite pentru a fi studiate sociologic).

Dar mai bine să începem prezentarea ariei de aplicare a metodei, căci atunci şi rolul şi importanţa pentru sociologie, dar şi specificul său în cadrul disciplinei noastre se vor lămuri, sper, pe deplin.

1. Tipuri de documente

Fără pretenţia de a da o definiţie riguroasă, vom înţelege prin document orice obiect material realizat de om, capabil să înmagazioneze o anumită informaţie. O precizare se impune însă imediat. Practic orice lucru construit sau chiar atins de mîna omului, ca să spun aşa, va reţine o informaţie despre cel care l-a realizat sau manipulat. O unealtă de muncă, o clădire, un drum, o simfonie, o armă, un poem etc. sînt cîteva exemple de entităţi create de om ce pot fi - şi sînt efectiv -folosite în reconstituirea vieţii sociale din epoca respectivă. Dar nu toate aceste documente au aceeaşi relevanţă pentru sociolog ca şi pentru istoric sau etnolog, chiar dacă nu este posibil să se traseze aprioric o linie despărţitoare în afara căreia să plasăm exact toate obiectele ce vor intra în sfera de preocupări a sociologului, cu titlul de documente. In principiu deci vom admite că orice astfel de entitate poate deveni un document sociologic, dar vom înţelege că, în realitate, sfera celor care sînt efectiv utilizate aici e mult mai redusă, sau că există o anume ierarhie, destul de greu, de altfel, de precizat şi ea, a utilităţii documentelor pentru sociolog.

In cele ce urmează voi da o clasificare a documentelor, apropiată de cea propusă de alţi autori, precum M.Duverger (1964) sau M.Grawitz (1972), rostul ei fiind doar acela de a introduce o anumită ordine sau de a constitui un fir conducător pentru o trecere în revistă cît mai completă a tipurilor de documente utilizabile.

1.1. Documente scrise de natură nesociologică

a) Statisticile oficiale. Sînt documente, cu caracter oficial, întocmite de unele organisme, pe baza numărării elementelor ce prezintă anumite caracteristici. Exemplul tipic de astfel de documente îl reprezintă rezultatele recensămintelor. Aceste operaţii se efectuează, la ora actzală, aproximativ din lo în lo ani, în mai toate ţările lumii, avînd drept scop furnizarea unor informaţii demo-sociale necesare pentru factorii de conducere în stat. In intervalele dintre recensăminte unele date - ca cele referitoare la mişcarea populaţiei, de exemplu - sînt reactualizate, an de an, pe baza înregistrărilor statistice curente. In cazul datelor demografice de bază, informaţiile pot fi ţinute la zi; cunoscînd structurile populaţiei la recensăminte şi evenimentele demografice din cursul unui an, se pot calcula relativ uşor volumul şi structura populaţiei în orice moment dorim.

La noi în ţară, statisticile curente anuale se publică în Anuarul statistic, care cuprinde, în cea mai mare parte, indicatori de factură economică, dar şi unele date demografice şi geografice de bază, Detaliile de prezentare coboară pînă la nivelul judeţelor şi/sau mediului urban şi rural. Pe lîngă lucrarea pomenită, Direcţia Centrală de Statistică mai publică, din cînd în cînd, culegeri de date statistice pe anumite domenii, editează Revista de statistică, unde pot fi găsite,de asemenea, informaţii statistice oficiale.

Importanţa surselor statistice pentru sociolog este extrem de mare, deoarece de aici el obţine date relativ sigure şi cu o arie de cuprindere maximă asupra societăţii în cauză. Informaţia aceasta primară, supusă unor prelucrări statistice adecvate şi culeasă pe o secvenţă temporală suficient de lungă, poate constitui un obiect perfect de analiză sociologică în sine sau poate folosi la conturarea unui cadru de referinţă necesar în evaluarea datelor culese prin alte metode. Din păcate însă, aceste calităţi sînt de multe ori contrabalansate de penuria informaţiilor oficiale care să intereseze în mod deosebit pe sociolog; aşa cum spuneam, statisticile oficiale conţin - şi la noi şi aiurea - foarte multe informaţii de natură îngust economică şi foarte puţine referitoare la alte aspecte sociale.

b) Arhive publice şi alte documente oficiale. Arhivele sînt colecţii de documente al căror conţinut poate ajuta sociologul să înţeleagă o serie de fenomene ale vieţii sociale; ele sînt evident mai folositoare atunci cînd dorim să cercetăm evoluţia sau geneza unor fenomene, deci cînd abordăm trecutul unor comunităţi umane. Este clar că arhivele sînt mult mai des frecventate de către istorici decît de sociologi, dar nu sînt ocolite nici de aceştia din urmă.

Un interes foarte mare pentru sociolog îl prezintă documentele oficiale ale statului ce sînt aduse la cunoştinţa publicului prin Buletinul Oficial, şi, cele mai importante, prin mijloacele de comunicare de masă . De o deosebită importanţă sînt la noi în ţară documentele Partidului Comunist Român, elaborate cu ocazia plenarelor Comitetului Central, a conferinţelor naţionale ale partidului, a congreselor. In aceste documente apar, pe lîngă date privind evoluţia unor indicatori socio-economici, şi analize aprofundate ale dezvoltării societăţii noastre, deci un întreg arsenal de concepte şi idei privind conducerea, organizarea şi evoluţia vieţii sociale în România.

c) Presa. Prin caracterul său de oglindă a faptelor cotidiene, de mai mică sau mai mare importanţă socială, presa este, în majoritatea ţărilor, un izvor de informaţii de o bogăţie inepuizabilă pentru cel care doreşte să descifreze prin intermediul ei anumite probleme ale unei comunităţi. Desigur, există diferenţe sensibile de la ţară la ţară în privinţa modului în care presa oglindeşte cu fidelitate faptele, dar oricîte obstacole ar ridica orientarea politică sau ideologică a diverselor ziare, pînă la urmă o analiză atentă poate să descrifeze esenţialul.

Caracteristic pentru presa din ţara noastră este faptul că ea reflectă, în cea mai mare parte a conţinutului său, aspecte majore ale societăţii româneşti de astăzi, putînd deci conduce sociologul spre cunoaşterea unor elemente de mare profunzime. Dar chiar realizată pe laturile mai puţin importante, cum ar fi anunţurile la mica publicitate, faptele diverse etc. analiza de conţinut a ziarelor nu este lipsită de interes.

d) Documentele instituţiilor, organizaţiilor, întreprinderilor etc. dau, în marea lor majoritate, informaţii despre modul de organizare, de funcţionare, despre numărul şi statutul persoanelor

ce iau parte la activitățile respective, despre relațiile cu alte unități și, în general, despre locul și rolul jucat de acestea în viața socială. Așa de exemplu, analiza proceselor verbale din ședințele C.O.M.-ului unei intreprinderi poate spune lucruri foarte interesante sociologului, chiar făcînd abstracție de conținutul tehnic al discuțiilor purtate aici - și mai ales făcînd abstracție de acesta. Se relevă astfel aspecte semnificative ale procesului de luare a deciziilor, gradul de participare la acest proces a reprezentanților muncitorilor, stilul de muncă al conducătorului intreprinderii etc. Nu mai vorbesc de faptul că documentele aflate la serviciul personal al unității sînt surse fără a căror consultare sociologul industrial nu poate practic studia o serie de probleme cum ar fi fluctuația forței de muncă și altele.

Exemplul cu intreprinderea, deși foarte sugestiv, este departe de a se încadra în rîndul celor mai semnificative cazuri, deoarece documentele ce se găsesc aici vizează un sector de activitate relativ restrîns și un volum de persoane modest ca mărime. Alte unități posedă documente cu deschidere mult mai amplă; mă gîndesc la valoroasa(pentru sociolog) documentație ce se găsește la diverse organe ca cele care dirijează, pe raza unui județ, de exemplu, activitatea școlară, sanitară, financiară, judecătorească etc. Cu siguranță că informațiile respective sînt folosite de organele de specialitate, dar sociologii le-ar putea valorifica din alte puncte de vedere, găsindu-le sensuri mai largi, descoperind corelații, determinări sau dependențe între fenomene ce ies în afara unei ramuri înguste de activitate. Din păcate, astfel de surse sînt rareori exploatate în cercetarea sociologică - și nu numai la noi - lucru care se datorează în principal faptului că accesul la astfel de documente nu este ușor.

e) <u>Documente personale</u>. Este vorba aici de scrisori, jurnale, autobiografii, certificate, diplome etc., deci de tot ceea ce poate da seama despre viața unei persoane sau a familiei sale, despre raporturile între indivizi sau micrigrupuri sociale, dar și de materiale ce conțin relatări despre fenomene cu o sferă mai largă, cu implicații sociale mai profunde.

Referitor la această sursă informațională, toate manualele

de metode pomenesc ca exemplu, ca model de analiză deosebit de reuşită celebra lucrare a sociologilor Thomas şi Znaniecki, Ţăranul polonez , apărută în 1919 în S.U.A. Studiind un număr impresionant de documente persoanale (scrisori, jurnale, autobiografii etc.), dar şi de altă natură (articole de ziare, arhive parohiale, rapoarte ale tribunalelor sau ale organismelor de ajutor social), autorii au urmărit, în special, să pună în evidenţă procesul de transformare a satului polonez ca urmare a schimbărilor survenite, la începutul secolului, în sistemul de valori şi atitudini al ţăranilor, schimbări accelerate şi de migraţia masivă în Statele Unite. Cei doi sociologi consideră că documentele persoņale, în speţă autobiografiile, constituie tipul perfect de material sociologic.

f) Literatura poate fi considerată fără nici o rezervă drept sursă de informaţii asupra vieţii sociale dintr-o anumită epocă, indiferent de interpretările şi concepţiile personale ale autorilor. Pentru a ne da seama de forţa pe care literatura o posedă în această privinţă este suficient să ne gîndim - desigur cei care nu sîntem istorici de meserie - la societatea rusă sau la cea franceză din secolul trecut şi, încercînd să decelăm sursele cunoştinţelor noastre despre ele, vom ajunge la concluzia că sîntem în proporţie covîrşitoare tributari imaginilor pe care marii romancieri ai celor două literaturi ni le-au transmis.

g) Alte lucrări tipărite precum dicţionarele, enciclopediile, afişele etc. sînt considerate, de asemenea, ca obiecte purtătoare de informaţie relevantă pentru sociolog.

1.2. Documente scrise sociologice

Voi cuprinde sub acest titlu toate documentele scrise elaborate în scop sociologic, în cadrul unor cercetări empirice. Ele pot fi de două feluri:

- materiale ce provin din cercetări anterioare celei în curs, cercetări efectuate fie de sociologul în cauză, fie de către alţii, şi care se găsesc sub formă de rapoarte de cercetare, articole cărţi, arhive sociologice etc. tratînd o problematică identică sau apropiată de cea urmărită;

- materiale constituite în cursul investigaţiei respective, ce se cer ele însele cercetate cu tehnici ale analizei documentelor (chestionare cu întrebări deschise, interviuri,etc.)

Cîteva precizări sînt necesare aici. De regulă, manualele nu includ asemenea documente în clasificările lor, fie probabil din cauza unor omisiuni involuntare, fie poate datorită temerilor de a nu extinde prea mult aria metodei, ceea ce ar face, la urma urmei, să-şi piardă specificitatea şi să poată fi invocată în orice gen de cercetare. Mai întîi , deci, trebuie să înlăturăm tocmai o asemenea posibilitate. Este clar, cum spuneam şi cu altă ocazie, că studiul bibliografiei este un punct important în pregătirea oricărei cercetări şi, prin urmare, dacă analiza documentelor rămîne o metodă specifică în cercetarea sociologică, ea nu trebuie să se suprapună cu acest studiu. E vorba în cazul metodei aici prezentate de o utilizare a <u>informaţiilor empirice</u> din alte cercetări sociologice, la fel cum astfel de informaţii sînt căutate în orice document, în timp ce prin studiul bibliografic urmărim cu precădere aspecte teoretice (precizarea noţiunilor, formularea ipotezelor etc.) şi doar parţial date empirice, şi acestea numai cu scopul de a găsi elemente de comparaţie pentru datele cercetării noastre iar nu ca fapte supuse la rîndul lor investigaţiei.

Reamintin că obiectul cursului acesta îl constituie prezentarea modalităţilor de cercetare sociologică <u>empirică</u>, vom observa totuşi că în cazul aplicării analizei documentelor la materiale sociologice, distincţia empiric-teoretic nu este foarte netă. Orice studiu sociologic pe care-l vom folosi ca sursă informaţională este, într-o proporţie sau în alta, un amestec de informaţii empirice şi constructe teoretice care au ghidat cercetarea în cauză. Ar fi absurd să se creadă că cele două elemente sînt detaşabile şi că s-ar putea extrage date empirice în formă pură. Dar aceasta nu ne împiedică totuşi să distingem cele două unghiuri principale din care sînt privite documentele socilogice: ca surse de informaţii despre realitatea studiată şi ca teorie sau metodologie de studiere a aceleiaşi realităţi.

Aşadar <u>analiza secundară a datelor sociologice</u> poate fi considerată în mod cu totul îndreptăţit ca o metodă (variantă a metodei analizei documentelor) importantă în arsenalul cercetării

sociologice. Mai mult, această metodă nu este destinată doar să acompanieze alte metode într-o investigaţie oarecare, ci poate constitui singură o cale eficientă de a obţine rezultate relevante şi importante, derivate din cercetări sociologice anterioare, dar nicidecum reductibile la concluziile acestor cercetări. Este limpede că dacă sociologul dispune de o masă cît mai mare de informaţii empirice (comparabile) despre un anume fenomen, el va avea şanse sporite de a decela unele regularităţi, conexiuni etc. care n-au putut fi puse în evidenţă de nici un material sociologic care i-a stat la bază. Un exemplu elocvent în această privinţă îl constituie lucrarea lui R.Boudon L'Inégalité des chances (1973), unde sociologul francez sintetizează rezultatele a zeci de cercetări asupra inegalităţilor de şanse în faţa şcolii şi creionează implicaţiile acestora asupra mobilităţii sociale.

Referitor la al doilea punct menţionat (analiza documentelor produse pe parcursul cercetării în cauză), se vede că aici metoda este utilizată doar în prelungirea altei modalităţi de culegere a datelor (ancheta sau interviul) şi nu în mod independent. Textele lungi ale unor întrebări deschise din chestionar sau ale interviurilor apar ca documente de sine stătătoare asupra cărora sociologul trebuie să se aplece cu instrumente de cercetare adecvate.

1.3. Alte surse

In afara documentelor scrise, există şi alte documente ce intră în cîmpul analizei sociologice; e drept însă că folosirea lor este incomparabil mai redusă în disciplina noastră decît în alte ştiinţe sociale învecinate: istoria, etnografia, antropologia etc.

Asemenea surse le-am putea clasifica, împreună cu Duverger (1964), în felul următor:

a) Obiecte tehnice (unelte, clădiri, mobilier etc.)

b) Documente iconografice (desene, gravuri, tablouri, sculpturi), fotografice şi cinematografice.

c) Documente fonetice (discuri, benzi de magnetofon, casete).

x

x x

In exemplificările de pînă aici am sugerat tot timpul o singură perspectivă din care se efectuează analiza documentelor, şi anume acea care vizează conţinutul documentelor ca reflectare, reproducere, înregistrare, adică drept sursă de informaţii despre anumite fapte sociale exterioare şi anterioare documentelor respective.

Se impune însă cu necesitate precizarea unui al doilea punct de vedere din care documentele îl interesează pe sociolog, aspect ce, în multe situaţii, poate să fie chiar mai important decît cel precedent. Se înţelege uşor că orice document este el însuşi un "obiect" social, la elaborarea căruia a participat un individ, un grup, o instituţie etc. Deci documentul este rezultatul unei activităţi umane, determinată de anumite fapte, împrejurări sociale, mobilizînd o serie de forţe omeneşti în realizarea lui şi care, odată realizat, mai cu seamă dacă posedă o largă circulaţie, va fi generator de efecte sociale.

Aşadar, o a doua perspectivă sociologică va fi cea care încearcă să dea răspunsuri la întrebări precum: cine a produs documentele, cu ce scop, cui i se adresează informaţia conţinută, ce gen de informaţie manipulează unele documente în comparaţie cu altele, ce efect social a produs elaborarea şi circulaţia unui document etc. In consecinţă, metoda analizei documentare cunoaşte două variante de bază: una pe care o voi numi, în lipsă de un termen probabil mai adecvat, directă, utilizează în mod nemijlocit informaţia conţinută în documente ca atare, deci ca informaţie despre fapte sociale ce nu au altă legătură cu documentele decît că acestea din urmă dau seama despre primele şi alta, să-i zicem indirectă, în care documentul însuşi are o semnificaţie socială, iar informaţia conţinută în el cere o interpretare deosebită, mediată de contextul social în care documentul a fost elaborat.

Practic, fiecare gen de documente este capabil să ne furnizeze informaţii indirecte; unele precum obiectele tehnice, sînt aproape în exclusivitate purtătoarele unei astfel de informaţie. Dar chiar şi cele mai importante documente, din punctul de vedere al bogăţiei de cunoştinţe directe despre societate, se pretează la fructuoase analize sociologice din unghiul al doilea menţionat. Astfel, presa în ţările occidentale constituie un element privilegiat de analiză documentară indirectă. Studiind, de exemplu, relatările pe care diferite jurnale le dau aceloraşi evenimente, se pot trage concluzii interesante despre orientarea ziarelor în cau-

ză, tot aşa cum analiza conţinutului unei publicaţii (genul rubricilor sale, extinderea acestora, natura valorilor promovate ş.a.m.d) este în măsură să ofere cunoştinţe asupra publicului căruia i se adresează, asupra ideologiei promovate, deci asupra intereselor de clasă sau grupele pe care le apără. Cercetările asupra conţinutului filmelor scot în evidenţă, de asemenea, o mulţime de lucruri importante, fie pe planul evoluţiei unor valori sau norme sociale, al tipului de eroi ori de conflicte promovate în cadrul unei cinematografii naţionale, fie diferenţele - după parametri similari - între producţiile cinematografice din două ţări la un moment dat. Natural că şi programele posturilor de radio sau televiziune se pretează din plin la acest gen de analize.

Un domeniu în care analiza de tip indirect a documentelor îşi găseşte un loc de seamă îl constituie sociologia politică. Discursurilor liderilor politici, de pildă, sînt studiate extrem de minuţios: apariţia unor teme sau cuvinte cheie noi ori repetarea unora mai vechi sînt indicatori principali după care se poate judeca orientarea politică a conducerii partidelor. Programele, platformele electorale, afişele sau lozincile partidelor nu conţin numai informaţii sociale generale - şi poate nu în primul rînd astfel de informaţii -, ci indicii asupra orientării, a poziţiei sociale a partidului respectiv.

Chiar şi lucrările de sociologie, atît empirice cît şi teoretice - ca, de altfel, orice alt gen de lucrări ştiinţifice - constituie obiect de analiză indirectă, fiind astfel purtătoare de informaţii pentru sociologia sociologiei (sau sociologia ştiinţei, mai general). Analiza conţinutului revistelor sociologice, după tematica abordată ori după metodele utilizate în cercetare, studierea întregului volum de publicaţii sociologice într-un interval de timp şi într-o ţară anume, analiza trimitelor bibliografice dontr-un set de lucrări, după naţionalitatea autorilor sau limba în care au scris, după anul apariţiei etc. sînt doar situaţiile cele mai frecvent întîlnite ce se încadrează perfect în problema aici descrisă.

Asemenea ilustrări ar putea continua pe mai multe pagini, dar exemplele precedente cred că sînt concludente pentru a reflec-

ta bogăţia de perspective din care documentele pot fi analizate sociologic, indiferent de natura sau destinaţia lor. Probabil că dihotomia mai sus menţionată (analiză documentară directă şi indirectă), ca orice dihotomie, este un pic rigidă şi ar fi nevoie de eforturi substanţiale de clasificare a tuturor unghiurilor de vedere, a tuturor pistelor ce-i sînt deschise sociologului în elaborarea documentelor ca obiect de analiză. Deşi idei similare pot fi găsite în multe cursuri şi tratate, în sensul că anumite genuri de documente se pretează la mai multe posibilităţi de analiză, nicăieri, personal, nu am găsit o abordare generală a analizei documentare după asemenea posibilităţi. Iar distincţia operată aici între o analiză directă (care merge pe linia folosirii informaţiei din documente ca atare, deci ca informaţie despre anumite fapte sociale petrecute)şi una indirectă (în care documentul este un fapt social sau conţinutul să nu interesează atît ca informaţie asupra evenimentului relatat ci, în primul rînd, asupra celui care relatează, a felului în care relatează, vizînd, prin urmare, o informaţie dincolo de cea aparentă) această distincţie, zic, chiar dacă poate prea abruptă,mi se pare totuşi esenţială.

De aici decurge şi o problemă metodologică importantă. In primul tip de analiză, cercetătorul trebuie să aibă mereu în vedere şi să verifice dacă informaţia conţinută în materialele ce urmează a fi investigate este corectă; el va exclude din capul locului din cîmpul de cercetare orice document cu un conţinut îndoielnic iar atunci cînd bănuieşte existenţa unor distorsiuni, în lipsă de alte surse mai adecvate, va încerca să le corecteze. Astfel, dacă - aşa cum se întîmplă frecvent în ţările occidentale - mai multe ziare dau relatări şi interpretări diferite aceluiaşi eveniment, prin compararea unei game largi de texte şi eliminarea distorsiunilor datorate orientării ziarelor, sociologul poate ajunge uneori la o informaţie relativ obiectivă.

In al doilea gen de analiză, asemenea probleme nu se pun deci conţinutul documentelor nu va fi judecat în termeni de adevărat sau fals, ceea ce face ca sfera de cuprindere să fie aici potenţial mai largă decît în cazul precedent. Singura problemă ce ţine de alegerea documentelor este aceea de a alege surse capabile să ofere un suport solid de cercetare, adecvat scopului ce şi l-a propus sociologul.

2. Tehnici de analiză

După rapida trecere în revistă a "obiectelor" de analiză şi a punctelor de vedere din care ele sînt interesante şi importante pentru investigaţia sociologică empirică, urmează în mod logic să vedem cum se realizează concret analiza, care sînt tehnicile şi instrumentele ataşate metodei. Din păcate un studiu sistematic în acest sens este greu de intreprins, căci puţine sînt elementele specifice muncii sociologului în procesul de studiere a documentelor; vreau să spun că activitatea lui în acest domeniu nu se deosebeşte din punct de vedere metodic de cea a istoricului a criticului literar sau a cercetătorilor din alte domenii.

Desigur, vorbind despre modalităţile de fructificare a informaţiei din documente, trebuie să detaşăm, în primul rînd, acele surse în care informaţia sociologică apare sub formă numerică, de regulă rezultat al numărării şi, mai rar, al măsurării unor caracteristici. Astfel de date intră imediat şi firesc în angrenajul clasic de analiză statistică, asupra căruia ne vom opri în partea a treia a cursului nostru. Voi menţiona aici doar faptul că sociologul trebuie să se manifeste extrem de prudent atunci cînd foloseşte date statistice culese de alte persoane. In principiu, utilizarea acestor cifre trebuie să fie precedată de cunoaşterea exactă a modului în care ele au fost constituite, adică a unităţilor statistice numărate şi/sau a însuşirilor măsurate, precum şi a gradului de precizie sau temeinicie cu care s-au realizat operaţiile de mai sus. Nu puţine sînt cazurile cînd, pornind de la cifre incorect constituite, se aplică proceduri statistice sau matematice foarte sofisticate ce nu pot să producă decît rezultate iluzorii. De asemenea, lipsa de unitate în definirea unităţilor de analiză face inutilizabile datele statistice de care sociologul dispune sau, şi mai rău, în cazul cînd asemenea inadvertenţe trec neobservate, comparaţiile şi concluziile vor fi, forţamente, false. A compara, de exemplu, nivelele de urbanizare în două ţări, fără a ţine seama de criteriile diferite după care o localitate primeşte statutul de oraş (se poate întîmpla ca în una să fie considerată urbană orice localitate ce depăşeşte 2000 de locuitori şi în alta pragul minim să fie urcat la 10.000), este evident un nonsens, oricît de precise ar fi statisticile ce ne stau la dispoziţie. Sau, un

alt exemplu în legătură cu precizia datelor. S-a observat că în unele ţări în curs de dezvoltare ratele de natabilitate au manifestat o tendinţă de creştere în ultimele decenii, opusă deci celei aşteptate. La o analiză atentă s-a dovedit că o serie de astfel de cifre se datorează pur şi simplu unei înregistrări mai corecte, pe măsura trecerii timpului, a cazurilor de naştere în aceste ţări.

Luînd la o parte documentele a căror analiză este transferabilă direct procedurilor statistice, vom observa că celelalte surse documentare pot fi supuse, în principiu, unor tehnici de analiză calitativă sau cantitativă. Tehnicicle calitative sînt, evident, cele mai curent întrebuinţate şi în sociologie ca şi în alte discipline socio-umane. Aici extragerea informaţiei din documente nu face apel la instrumente deosebite, ci se folosesc cele clasice şi banale: fişa, conspectul, rezumatul etc. Iar analiza urmează, de asemenea, căile cunoscute: remarcarea apariţiei unor informaţii, comparaţii privind conţinutul documentelor diverse, detaşarea ideilor principale într-un text, reconstituirea contextului social-istoric în care s-a elaborat documentul etc., fiecare cercetător(sociolog, istoric sau critic literar) efectuînd-o fără să-şi conştientizeze în mod deosebit actele, aşa cum - după remarca lui M.Duverger - domnul Jourdain făcea proză fără să ştie.

Este normal că, în lipsa unor cadre obiective de analiză, oricît de conştiincios, informat sau bine intenţionat ar fi un cercetător, rezultatele muncii sale nu sînt scutite de influenţe subiective, de interpretări ale textelor determinate de condiţia socială, morală sau intelectuală a celui care întreprinde analiza. Acest lucru este resimţit în toate domeniile care folosesc analiza documentelor, iar ultima jumătate de secol este caracterizată de eforturi multiple în depistarea acelor tehnici care să permită, în măsura posibilului, depăşirea impresionismului (deci a subiectivismului) în astfel de analize. Una dintre direcţiile pe care s-a mers - şi, repet, nu numai în sociologie - este cea a analizei de tip cantitativ, al cărei obiectiv fundamental îl constituie transformarea informaţiei descriptiv-discursive a documentelor în unităţi (entităţi) numărabile, sau , eventual, măsurabile, după anumite însuşiri. Tehnica impusă şi perfecţionată de sociologie şi psihologie socială capabilă să satisfacă un asemenea deziderat poartă denumirea de analiză de conţinut , iar obiectul rîndurilor

ce urmează îl constituie schiţarea liniilor generale de punere în lucru a acestei tehnici.

Mai întîi însă iarăşi o remarcă. In contextul actual, caracterizat de un contraatac destul de viguros la adresa pozitivismului din partea unor orientări filosofice opuse, şi în sociologie, în cadrul mai larg al frontului anticantitativist, se pot întîlni frecvent idei care neagă orice valoare studiilor de tip cantitativ în analiza documentară, făcînd apologia tehnicilor comprehensive, hermeneuticii etc. Este clar că şi aici, ca şi referitor la multe alte probleme, astfel de reacţii sînt de înţeles iar critica apare justificată atunci cînd vizează excesele cantitativiste; dar nu o dată criticile sînt generale, fără discernămînt şi de aceea nocive. De asemenea, este iarăşi limpede că o analiză strict cantitativă a documentelor este practic imposibilă şi că în cadrul analizei de conţinut, de pildă, aspectele cantitative şi cele calitative se împletesc poate mai strîns decît în orice altă metodă sau tehnică de cercetare sociologică.

Revenind la analiza de conţinut, vom accepta, pentru moment, în loc de definiţie, următoarea caracterizare: ea este o tehnică de cercetare a documentelor ce furnizează o descriere obiectivă sistematică şi cantitativă a conţinutului acestora, urmînd ca aspectele concrete ale tehnicii să fie înţelese din descrierea ce urmează.

Nu voi intra în detalii privind istoricul problemei. Deşi unele încercări de tratare cantitativă a textelor sînt cunoscute de multă vreme, lucrările de metodologie înclină, în majoritate, să-l considere drept părinte adevărat al analizei de conţinut pe Harold D.Lasswell, un psihosociolog american, care a elaborat principiile analizei şi a utilizat-o împreună cu colaboratorii, în domeniul sociologiei politice, în perioada celui de al doilea război mondial. Pornită deci iniţial de la studiul comunicaţiilor politice (discursuri, comunicate oficiale, ştiri din presa scrisă şi vorbită etc.), analiza de conţinut s-a dovedit a fi o tehnică ce se poate aplica şi în alte domenii cu aceeaşi eficienţă.

In principal aria de utilizare a tehnicii o reprezintă documentele scrise, iar aspectul cantitativ despre care aminteam este cel statistic (de numărare). Din acest motiv, pentru a fi posibilă aplicarea tehnicii este nevoie ca "populaţia" documentelor

ce intră în studiu să fie suficient de largă, altfel nu se pot pune în evidenţă regularităţile statistice. In cazul în care volumul documentelor ajunge să depăşească posibilităţile unei analize exhaustive, se pot aplica şi aici procedee de eşantionare, alegîndu-se un eşantion de texte care să posede o reprezentativitate corespunzătoare necesităţilor cercetării. De exemplu, dacă lotul integral de documente este repezentat de numerele unui ziar, cotidian, extinse pe durata unui deceniu, vor trebui cercetate peste 3.1oo de numere (presupunînd că o zi pe săptămînă ziarul nu apare), ceea ce, exceptînd cazul unei analize foarte superficiale, e greu de realizat. Un eşantion de cca 5oo de ziare (unul pe săptămînă) poate fi suficient pentru atingerea scopurilor urmărite.

Stabilirea populaţiei şi a eşantionului de documente se face în corelaţie strînsă cu precizarea foarte exactă a problemei de studiat. In mod deosebit, în acest gen de analiză problema cercetată trebuie să primească o delimitare riguroasă şi să fi transpusă corect şi complet în indicatori empirici. Chestiunea centrală a studiului se concretizează, de regulă, în una sau mai multe caracteristici, a căror prefigurare cantitativă se urmăreşte în studiile mai simple (explorative); acestora li se adaugă şi alte caracteristici, auxiliare, cu rol de variabile explicative, în studiile mai complexe.

De exemplu, un studiu explorativ poate avea drept scop evidenţierea "tematicii articolelor de fond din ziarul Scînteia" pe o perioadă dată. Expresia dintre ghilimele este atunci singura caracteristică ce urmează a fi construită în urma studiului empiric. La fel şi în analize cu teme ca: "valorile morale promovate de emisiunile de tineret ale televiziunii", "tipurile socio-profesionale de personaje din reportajele Scînteii tineretului", "ocupaţia eroilor principali din romanul românesc din ultimii 35 de ani" etc.

Adîncirea analizei poate merge cel puţin pe două coordonate principale: fie o aceeaşi caracteristică se va studia pe subpopulaţii distincte de documente (de pildă, prima temă menţionată va fi urmărită pe anumite perioade de apariţie a ziarului Scînteia ori caracteristica va fi elaborată pentru ziare diferite etc.), fie populaţia rămîne aceeaşi şi se construiesc mai multe caracte-

risticî pe ea, urmînd ca acestea să fie apoi studiate prin proceduri statistice precum asocierea sau corelaţia. A doua variantă se apropie de cercetările pe bază de anchetă, numai că dacă printr-un chestionar putem ajunge să obţinem zeci sau chiar sute de caracteristici ale populaţiei, în analiza de conţinut cu greu se vor obţine mai mult de 3-4 asemenea indicatori, din cauză că, pe de o parte, documentele nu sînt capabile să "răspundă" la atîtea întrebări ca oamenii iar, pe de altă parte, se ridică serioase dificultăţi practice în construcţia simultană a unui număr ridicat de caracteristici pe aceleaşi documente.

Se înţelege uşor că mulţimea, mai largă sau mai restrînsă, a caracteristicilor propuse spre construcţie este dependentă de conţinutul informaţional al documentelor, deci însăşi problema de cercetat trebuie corelată corect cu sursa informaţională. De asemenea, populaţia documentelor se cere circumscrisă cel puţin cu tot atîta atenţie ca şi populaţia umană în cadrul anchetelor. In fine, atunci cînd este vorba de documente ce apar într-o succesiune regulată, cum sînt de pildă ziarele, trebuie avut în vedere ca perioada analizată să fie semnificativă pentru faptele sociale urmărite. Spre exemplu, dacă ne va interesa felul în care <u>Scînteia tineretului</u> reflectă participarea tinerilor la campaniile agricole de recoltare, atunci se cer alese în fiecare an numere din perioadele cînd se realizează aceste campanii. Desigur, cazul acesta este banal, dar el reflectă o situaţie mai generală, şi anume aceea că din cauza periodicităţii unor acţiuni sociale, trebuie acordată toată atenţia felului în care se aleg documentele.

In continuare să presupunem, pentru simplificarea expunerii, că analiza de conţinut ce urmează să o intreprindem vizează stabilirea unei singure caracteristici. După delimitarea exactă a populaţiei de documente şi precizarea semnificaţiei caracteristicii în cauză, urmează operaţia de stabilire a <u>categoriilor</u> caracteristicii.

Categoriile sau clasele (voi folosi mai ales primul termen, încetăţenit în literatura despre analiza de conţinut) nu sînt altceva decît rubricile în care conţinutul informaţiei documentare este clasat pe baza unui criteriu. Altfel spus, categoriile sînt omoloagele variantelor de răspuns la o întrebare din chestionar, numai că de regulă, problemele cercetate prin analiza de conţinut

sînt de asemenea natură încît evantaiul variantelor este foarte complex. Din această cauză, se consideră că operaţia esenţială în realizarea unei analize de conţinut, de care depinde în cea mai mare măsură succesul cercetării, o constituie stabilirea categoriilor. Practic, lista categoriilor unei caracteristici este instrumentul tehnicii de faţă şi faptul că unii autori îi ataşează aceeaşi importanţă ca şi chestionarului în anchetă este pe deplin justificat. Cum operaţia de stabilire a categoriilor este uneori foarte dificilă, ea nu poate fi pe deplin realizată totdeauna înainte de începerea cercetării. Asta înseamnă că în anumite situaţii evantaiul categoriilor se precizează, se constituie pe deplin doar pe măsură ce documentele sînt despuiate sau într-o fază pilot a cercetării, pe un subeşantion de texte.

Există, evident, foarte multe tipuri de categorii şi nu se pot elibera reţete valabile pentru orice cercetare, tot aşa cum nu se pot elabora chestionare universale. Voi menţiona cîteva tipuri cu caracter pur ilustrativ:

(i) Dacă se studiează conţinutul unor ziare, reviste, emisiuni de radio sau televiziune etc., categoriile vor suprinde subiectele tratate sau genurile de rubrici. De pildă, pentru programele de televiziune: filme, seriale, piese de teatru, emisiuni muzicale, varietăţi, emisiuni de ştiri etc. Intr-o eventuală analiză a revistei Viitorul social, articolele se pot clasifica după subramura sociologică în care se încadrează conţinutul lor ş.a.m.d.

(ii) Un alt tip de categorii deseori folosit este cel de valori. Astfel, pentru a realiza, de exemplu, o analiză de conţinut a manualelor de limba română din ciclul primar, ne putem propune drept scop să evidenţiem valorile promovate în lecţiile respective. Analize similare sînt imaginabile asupra unui set de opere literare, emisiuni radio sau de televiziune, articole de ziar etc.

(iii) O serie specifică de categorii rezultă din abordarea însuşirilor unor personaje: vîrsta, sexul, statului social, trăsăturile fizice sau de caracter etc. In parte, aceste categorii coincid, după cum se vede, cu întrebările de clasificare din chestionar.

(iv) In fine, un alt exemplu tipic îl reprezintă categoriile de apreciere. Ele corespund variantelor sub formă de scală ordinală ale întrebărilor de opinie din chestionar. Caracteristici

cu asemenea categorii se pot construi doar atunci cînd în documente se găsesc aprecieri suficient de clare asupra unor fenomene sociale. De pildă, dacă ziarele iau atitudine față de un eveniment politic, se va încerca să se încadreze aceste poziții în categorii de genul: aprobare totală, aprobare parțială (moderată), indecizie, dezaprobare parțială (moderată) , dezaprobare totală. Dat fiind faptul că textele nu sînt scrise ca să ne faciliteze nouă analiza de conținut, de cele mai multe ori este greu să realizăm un evantai prea larg de categorii de apreciere și ne vom mulțumi cu situația cea mai simplă: trihotomia favorabil , neutru (indecis), nefavorabil, termenii de apreciere putînd, evident, varia în funcție de problemă.

Pentru a-și îndeplini rolul important de le revine în sistematizarea materialului analizat, categoriile trebuie să fie:

- exhaustive (capabile să surprindă toate variantele posibile de apariție ale caracteristicii),
- exclusive (o unitate înregistrată într-o categorie să nu mai fie și în alta),
- obiective (să permită o clasificare a materialului care să depindă cît mai puțin posibil de subiectul care o realizează),
- pertinente (adică adecvate obiectivului urmărit de cercetare și conținutului documentelor).

Caracterul de exclusivitate nu este totuși general, așa cum nici variantele de răspuns la întrebările chestionarelor nu sînt totdeauna exclusive. Așa de exemplu, dacă efectuăm pomenita analiză a manualelor de citire, în cadrul unei lecții pot fi promovate mai multe valori, deci unitatea va trebui încadrată simultan în mai multe categorii.

Si cu aceasta am ajuns la a doua problemă fundamentală a analizei de conținut, și anume aceea de a preciza ce anume se urmărește în text, adică de a se specifica indicatorul empiric al categoriilor caracteristicii și de a se delimita spațiul unde acesta trebuie căutat. In cazul anchetei asemenea probleme nu se pun deoarece răspunsurile la întrebările chestionarului se adaptează se dau în conformitate cu aceste întrebări. In analiza documentelor, dimpotrivă, întrebările se adaptează la informația conținută în documente și cum există o mare variabilitate a textelor trebuie să știm cum să depistăm existența unui răspuns la o proble-

mă ce ne interesează. Apoi, în anchetă fiecare răspuns vine de la un individ uman, individ care se poate distinge net de ceilalţi, ca entitate. In cazul documentelor lucrurile nu sînt nici în această privinţă atît de simple. In exemplele precedente, am sugerat cîteva unităţi cum ar fi numărul unei publicaţii, o emisiune de radio sau de televiziune, o lecţie din manual, un articol dintr-o revistă ştiinţifică etc. Dar, pe de o parte, nu totdeauna documentele apar atît de clar fragmentate sau individualizate iar, pe de alta, fiecare unitate ca cele de mai sus poate fi descompusă în componente care, la rîndul lor, sînt apte de a fi luate în considerare ca unităţi distincte.

Va trebui deci stabilită, mai întîi, aşa-numita <u>unitate de înregistrare</u>, adică entitatea purtătoare a informaţiei căutate. Cea mai simplă unitate de înregistrare este, desigur, <u>cuvîntul</u>. Numai că limbajul natural este atît de complex încît - exceptînd situaţiile în care studiul este de natură pur lingvistică, ceea ce evident că pe noi nu ne interesează aici - rareori putem folosi acest indicator, din lipsa unei suprapuneri cît de cît acceptabile între cuvînt şi <u>noţiune</u>. Vreau să spun că, de regulă, în sociologie nu ne interesează cuvîntul ca atare, ci semnificaţia acestuia şi că în textele analizate putem întîlni cuvinte diferite cu semnificaţii identice sau acelaşi cuvînt, în alt context, cu semnificaţii diferite. Astfel, dacă în studiul unui discurs politic se urmăreşte apariţia indicatorilor pentru categoriile de "patriotism" şi "internaţionalism", pentru prima dintre ele se pot lua ca unităţi de înregistrare termen ca: patrie, popor, naţiune etc.

Se observă chiar şi din acest singur exemplu cît de dificilă este alegerea cuvintelor capabile să reprezinte indicatori siguri pentru categoriile urmărite. Din acest motiv, studiile sociologie bazate pe cuvînt ca unitate de înregistrare nu sînt cele mai frecvente. Doar atunci cînd se pot identifica unele cuvinte cheie, cum este cazul în analiza textelor de propagandă politică, şansele de a duce la bun sfîrşit o lucrare sînt mai consistente.

Unitatea de înregistrare cea mai des folosită este <u>tema</u>, adică acel fragment din text care, prin ideile conţinute, corespunde semnificaţiei uneia sau alteia dintre categoriile stabilite. Spre pildă, dacă în una din lecţiile de citire, ca să reiau un exemplu anterior, vom găsi descrierea unei bătălii pe care poporul român a

dus-o împotriva cotropitorului străin, atunci cu siguranţă că unitatea respectivă trebuie înregistrată în categoria "patriotism", din evantaiul claselor de valori presupuse a fi promovate în textele respective. Se vede că nici depistarea ideilor, ca unităţi de înregistrare, nu este un lucru chiar atît de simplu, ba uneori operaţia se dovedeşte chiar mai dificilă decît în cazul cuvintelor sau noţiunilor. Reperarea unei teme într-un text este un act puternic influenţat de trăsăturile celui care-l realizează şi, din acest motiv, apreciez că tehnica analizei de conţinut -chiar dacă sînt foarte riguros stabilite categoriile - nu poate fi realizată la cote calitative superioare decît de persoane cu multă experienţă şi cu un simţ analitic foarte dezvoltat.

După ce sociologul stabileşte cît mai concret posibil unităţile de înregistrare, adică indicatorii pe care trebuie să-i caute în text, urmează imediat precizarea ariei, a spaţiului în care respectivii indicatori vor fi căutaţi. Cu alte cuvinte, universul documentelor trebuie împărţit în fragmente bine individualizate, astfel încît pe fiecare asemenea fragment să se urmărească dacă tema sau cuvîntul apare sau nu. Aceste unităţi poartă numele de unităţi de reperaj . Pe lîngă exemplele de unităţi de reperaj deja date (emisiunea de radio sau televiziune, număr sau articol de publicaţie, lecţie de manual) se pot imagina încă multe altele. De pildă, dacă ne referim la un ziar şi dorim să luăm o unitate de reperaj mai mică decît numărul sau articolul, putem merge cu fragmentarea textului în coloane, fracţiuni de coloane, alineate sau chiar fraze. La programele de radio sau televiziune, unităţile de reperaj mai mici decît emisiunea sînt, de regulă, de natură temporală (număr de minute de emisie). Astfel, în cadrul unor programe radiofonice de profil literar, cronometrînd intervale de 2,3 sau 5 minute, putem urmări pe acestea dacă numele unui scriitor este sau nu pronunţat, dacă sînt emise sau nu anumite idei etc.

In sfîrşit, dat fiind faptul că unităţile de reperaj nu sînt totdeauna egale ca întindere, în anumite situaţii va trebui să se ţină seama şi de dimensiunile acestora. Analizînd, de pildă, numerele pe ultimii 2o de ani ai Revistei de filosofie şi alegînd articolul ca unitate de reperaj, este clar că apariţia unei teme într-o recenzie de o pagină nu poate fi identificată decît într-un mod foarte grosolan cu apariţia aceleiaşi teme tratată însă într-un

studiu de, să zicem,15 pagini. Prin urmare, aici se impune luarea în calcul şi a dimensiunii unităţii de reperaj sau a întinderii pe care tema o prezintă. Pentru aceasta se introduce încă o noţiune: unitatea de numărare ca subdiviziune minimă pentru unitatea de înregistrare. In cazul anterior, unitatea de numărare ar putea fi pagina (pentru a nu coborî mai mult) şi atunci fiecare temă ce apare într-un articol va fi luată în considerare de atîtea ori cîte pagini are articolul respectiv.

Evident că în practică lucrurile sînt ceva mai complicate. Rareori o temă se întinde pe întreaga arie a unităţii de înregistrare şi, din acest motiv, ea nu poate fi automat ponderată cu numărul unităţilor de numărare. Este posibil ca un studiu de 2o de pagini , luat ca unitate de înregistrare, să trateze o temă pe una sau două pagini, ceea ce înseamnă că tema respectivă va trebui numărată doar de una sau două ori şi nu de 2o. Sau, mai complicat, o temă poate apare pe mai multe pagini dar tratată de fiecare dată foarte sumar, într-o frază sau două (deci e departe de a ocupa întregul spaţiu al paginilor respective). In astfel de situaţii, se va decide de la început asupra modului de lucru: sau se va lua ca pondere numărul concret de pagini pe care tema e identificată , indiferent de întinderea ei, sau se va evalua în pagini textul rezultat din cumularea fragmentelor specifice temei.

Pentru analiza de conţinut a ziarelor, o unitate de numărare frecvent utilizată este cea cu care se poate măsura suprafaţa pe care se întinde textul, de exemplu centimetrul pătrat. Se ajunge astfel ca fiecare apariţie de temă să fie evaluată în cm^2, ceea ce înseamnă practic o trecere de la numărare la măsurarea extinderii temei.

Modul concret de realizare a unei analize de conţinut variază de la investigator la investigator şi în funcţie de natura documentelor cercetate, Atunci cînd avem de a face cu texte scrise - cazul cel mai frecvent = este recomandabil ca apariţia indicatorilor categoriilor să fie notată pe marginea textului cu simbolurile alese pentru categorii. In felul acesta se poate verifica mai uşor că fiecare unitate de înregistrare a fost analizată, se facilitează construcţia tabelelor de asociere între mai multe caracteristici şi se poate mult mai simplu repeta analiza, deci verifica rezultatele obţinute. Un exemplu cu două caracteristici e

uşor de imaginat în legătură cu analiza manualelor de citire despre care am mai pomenit. Fie prima caracteristică "valorile promovate", pe care o notăm cu A iar categoriile ei (valorile ce le vom distinge) cu $A_1, A_2, A_3, \ldots A_s$ şi a doua caracteristică "tipologia personajelor principale", notată cu B, avînd drept categorii (tipuri de personaje) $B_1, B_2, B_3, \ldots, B_t$. Dacă luăm lecţia drept unitate de înregistrare şi numărare, în acelaşi timp, înseamnă că trebuie să urmărim pentru fiecare lecţie dacă şi care dintre categoriile de valori şi personaje apar, notînd asemenea apariţii cu simboluri A_i şi /sau B_j. Astfel în dreptul unei lecţii putem avea următoarea notaţie: A_2, A_5, A_7 , B_3, corespunzînd faptului că am identificat aici trei categorii de valori iar personajul principal l-am clasat, după B, în categoria B_3. E posibil, evident, ca în dreptul unei lecţii să nu putem trece nici un simbol pentru o caracteristică sau alta. De asemenea, se cere subliniat că atunci cînd un simbol de categorie este marcat în dreptul unei lecţii el nu va mai fi marcat a doua oară chiar dacă ideia se repetă căci am ales lecţia, şi nu un fragment al acesteia, ca unitate de înregistrare şi numărare.

Rezultatul primar al analizei de conţinut îl constituie construcţia unor caracteristici statistice, de regulă calitative, adică a unor tabele simple de forma:

Categorii	A_1	A_2	...	A_s	Total
Nr. de apariţii	k_1	k_2	...	k_s	n

ce se pretează la toate operaţiile statistice ce vor fi descrise în partea a treia a cursului de faţă. O singură problemă specială se ridică aici, şi anume aceea că numărul de apariţii, <u>n</u> , nu coincide neapărat cu numărul unităţilor de înregistrare sau numărare, deoarece nu totodeauna categoriile sînt exclusive şi, pe de altă parte, există unităţi în care nu se poate identifica niciuna din categoriile caracteristicii (situaţia analoagă, formal, nonrăspunsului la o întrebare din chestionar). De aceea, cifrele relative, în speţă procentele, se calculează la mai multe totaluri: al numărului de apariţii, al numărului de unităţi de înregistrare etc.

De pildă, dacă setul de manuale analizate cuprinde loo de lecţii, numărul de apariţii al valorii "patriotism" este 5o (această valoare e promovată în 5o de lecţii) iar numărul total de apariţii pentru întreg setul de valori este, să zicem 25o , atunci frecvenţa 5o se poate raporta fie la totalul de loo (5o% din lecţii promovează valoarea respectivă), fie la 25o (2o% dintre apariţii revin valorii "patriotism).

Am ales mai sus spre exemplificare nişte cazuri şi cifre fictive deoarece exemple de analiză de conţinut în literatura noastră sociologică sînt puţine şi nu foarte semnificative pentru scopurile didactice aici urmărite. Aş aminti totuşi cartea _Sociologie românească contemporană_ (1983) , în care sînt expuse, printre altele, şi rezultatele unei ample cercetări a scrierilor sociologice din perioada postbelică, cercetare ce poate fi asimilată analizei de conţinut. Unităţile de înregistrare au fost lucrările sociologice, dar "calibrul" lor foarte diferit i-a condus pe cercetători să grupeze materialul pe 6 dimensiuni: volume, culegeri de studii, articole, cercetări, teze de doctorat şi manuale. Caracteristica utilizată a fost _tematica_ lucrărilor, clasificată în 21 de categorii. Tabelul frecvenţelor, după care s-a întocmit şi ierarhia categoriilor, apare la pagina lo3 a lucrării citate.

Cap.VI. EXPERIMENTUL

In cadrul activităţii de cercetare, atunci cînd este vorba de ştiinţele empirice şi, mai ales, de cele ale naturii, experimentul este considerat, de regulă, ca fiind metoda cea mai consistentă de cunoaştere. Această poziţie privilegiată a experimentului în rîndul metodelor prin care omul reuşeşte să explice fenomenele naturii, prin care se descoperă , se evidenţiază şi se confirmă legăturile cauzale dintre fenomene, nu putea să nu atragă atenţia şi cercetătorilor din domeniul ştiinţelor sociale.

Dar specificul acestor discipline - şi al sociologiei, mai cu seamă - ridică de la bun început o serie de obstacole în calea utilizării metodei experimentale, provocînd astfel numeroase discuţii asupra posibilităţii însăşi a experimentării în ştiinţele umane, discuţii care umplu mii de pagini în lucrările de natură metodologică şi epistemologică. Dacă cineva şi-ar da silinţa, probabil că ar descoperi inimaginabil de multe poziţii , caracterizate prin nuanţe şi subtilităţi deosebite, care merg pe o scală ce începe cu negarea oricărei posibilităţi de experimentare în sociologie şi se sfîrşeşte cu concepţia care pune un semn de egalitate (în ceea ce priveşte utilitatea, posibilităţile de cunoaştere şi de realizare) între experimentul sociologic şi cel din ştiinţele fizice.

Chiar dacă lăsăm la o parte de la bun început filozofii şi sociologii care consideră realitatea socială complet distinctă (şi chiar opusă) faţă de cea fizică, cel puţin din punctul de vedere al posibilităţilor de cunoaştere, deci pe cei care introduc o ruptură între metodele de cunoaştere din ştiinţele naturii şi cele socio-umane, lăsînt, prin urmare, la o parte pe acei care, în mod normal, prin poziţia lor filozofică, nu pot accepta experimentul ca un mijloc de cunoaştere a vieţii sociale şi referindu-ne numai la cei care acceptă , implicit sau explicit, o unitate a metodelor de cunoaştere ştiinţifică, vom găsi, referitor la experiment poziţiile extreme amintite mai sus. Restrîngînd şi mai mult aria, şi făcînd apel numai la gînditori marxişti, vom identifica practic aceeaşi diversitate de poziţii . Astfel T.Nicola (1973,pp.58-59)

examinînd cîteva opinii exprimate într-o dezbatere din literatura sovietică, identifică trei puncte de vedere:

1. Experimentul este posibil numai în cunoaşterea fenomenelor naturii şi alte tehnicii.

2. Experimentul poate fi utilizat în studierea fenomenelor sociale, dar metoda este esenţialmente diferită de cea din ştiinţele naturii.

3. Experimentul se poate folosi cu succes în cercetarea fenomenelor sociale, adecvînd metoda la specificul acestora.

Deşi mărturisesc că nu-mi este prea clară distincţia dintre poziţiile 2 şi 3, bănuiesc totuşi că este vorba de o gradaţie a atitudinilor faţă de experimentul social.

Pentru a nu mai zăbovi prea mult în discutarea acestor poziţii, voi mai ilustra doar una ce mi se pare deosebit de semnificativă. Este vorba de concepţia lui E. Durkheim, sociologul care a militat cu atîta forţă pentru a impune necesitatea privirii faptului social ca un lucru, deci pentru a apropia cunoaşterea sociologică de cea din ştiinţele naturii. Imi permit să dau un citat mai amplu, de fapt începutul capitolului VI din Regulile metodei sociologice (1974) spre a înţelege exact poziţia lui în problema ce o discutăm aici (sublinierile îmi parţin):

"N-avem decît un mijloc pentru a demonstra că un fenomen este cauza altuia, acela de a compara cazurile în care ele sînt simultan prezente sau absente şi de a căuta dacă variaţiile pe care le prezintă, în diverse combinări de circumstanţe, dovedesc că unul depinde de celălalt. Cînd ele pot fi produse în mod artificial după placul observatorului, atunci metoda este experimentală propriu-zisă. Cînd, dimpotrivă, producerea faptelor nu este la dispoziţia noastră şi nu putem decît să le aproximăm aşa cum s-au produs spontan, atunci metoda care se întrebuinţează este aceea a experimentării indirecte sau metoda comparativă.

Am văzut că explicaţia sociologică constă exclusiv în a stabili raporturi de cauzalitate, fie că e vorba de a rataşa un fenomen la cauza lui sau, dimpotrivă, o cauză la efectele sale utile. Pentru că, pe de altă parte, fenomenele sociale scapă evident acţiunii operatorului, metoda comparativă este singura care se potriveşte sociologiei" (p.164)

Aşadar, este limpede, cred, deşi Durkheim mai foloseşte uneori în continuare termenul de experiment ca sinonim pentru metoda comparativă, că pentru sociologul francez metoda experimentală "propriu-zisă" este apanajul exclusiv al ştiinţelor naturii.

Problema care se ridică şi consider că trebuie rezolvată în acest moment este cea de a explica posibilitatea unei diversităţi atît de mari de opinii, care, cum spuneam, ating poziţii extreme. După părerea mea lucrurile se explică relativ simplu, în funcţie de două elemente esenţiale:

1. ce definiţie adoptăm pentru sociologie;
2. ce definiţie adoptăm pentru experimentul în sociologie.

Referitor la prima problemă, se poate uşor constata că dacă se ia pentru sociologie un obiect mai larg, în sensul în care unele definiţii vorbesc despre "societatea în ansamblul său", "sistemul social global" etc., atunci este firesc că posibilitatea experimentării este substanţial redusă. Dacă, dimpotrivă, se acceptă ca aparţinînd sociologiei fenomene sau procese ce caracterizează grupuri mai restrînse sau acţiuni sociale foarte particulare, atunci este iarăşi evident că o serie de obstacole "materiale" în calea experimentării cad.

In al doilea rînd, aşa cum vom vedea în continuare, dacă se identifică experimentul social cu modelul ideal de experiment din ştiinţele naturii, atunci şansele de a experimenta în sociologie - şi chiar în alte ştiinţe sociale - sînt mici. Dimpotrivă pe măsură ce se renunţă la condiţiile de definiţie a experimentului, deci dacă se acceptă definiţii mai puţin "tari", atunci posibilităţile de a realiza asemenea experimente cresc.

Avînd în vedere acest din urmă aspect şi pentru a putea oferi un ghid folositor pentru a înţelege multitudinea de atitudini referitoare la experiment, am considerat necesar să expun pe scurt esenţa metodei experimentale.

1. Ce este experimentul ?

Posibilităţile de definire a experimentului sînt extrem de numeroase, fapt care influenţează, cum spuneam şi atitudinile faţă de această metodă. Psihosociologul S.Chelcea (1982, p.95) aminteşte că E.Greenwood, căruia îi datorăm o lucrare fundamentală, Experi-

mental Sociology. A study in Method, apărută în 1945, identifică în literatura sociologică şi psihosociologică din primele patru decenii ale secolului nostru peste 1oo de definiţii ale experimentului.

In ceea ce mă priveşte, voi încerca să reliefez esenţa metodei experimentale, aşa cum ea s-a constituit în ştiinţele naturii, reţinînd acele caracteristici fundamentale pe care trebuie să le îndeplinească orice experiment indiferent de domeniul său de realizare.

Prima observaţie care se cuvine făcută este aceea că experimentul este superior observaţiei ca eficienţă cognitivă pentru că el implică toate rigorile acesteia, adăugînd şi altele suplimentare. Mă gîndesc în special la faptul că, la fel cu observaţia, experimentul presupune urmărirea conştientă şi sistematică a manifestării fenomenelor studiate, precum şi înregistrarea manifestărilor respective. Această urmărire poate fi executată cu (iar înregistearea facilitată de) diferite instrumente şi aparate, precum microscopul , ampermetrul etc. Mi se pare foarte limpede că un cercetător din cîmpul ştiinţelor naturii include asemenea activităţi în orice situaţie experimentală.

Ce aduce în plus experimentul ? După părerea mea, două lucruri, ambele la fel de importante şi deci indispensabile:

(i) Intîi, experimentul presupune, spre deosebire de observaţie, intervenţia activă a subiectului cunoscător în desfăşurarea fenomeneului studiat. Această intervenţie activă se manifestă fie prin introducerea, provocarea unui factor nou (unor factori noi), inexistent (inexistenţi) în realitatea studiată, fie prin dirijarea variaţiei unui factor existent (unor factori existenţi).

(ii) In al doilea rînd, experimentul presupune controlul unui număr de factori. Termenul de control trebuie înţeles aici într-un sens ceva mai restrîns decît îl are în limba engleză. Intr-adevăr, se poate înţelege prin control însăşi acea intervenţie de introducere sau variaţie a unui factor. Pentru a deosebi aspectele prezentate la (i) şi (ii), prin control voi înţelege numai o acţiune limitativă, de îngrădire a manifestării altor factori, cu scopul de a fi siguri că singura schimbare esenţială petrecută în complexul studiat este cea provocată de cercetător.

Prin urmare, experimentul presupune o intervenţie a subiectului cercetător în dublu sens: de producere, de provocare a unor manifestări şi de constrîngere, de limitare a altora. De exemplu, prin control în fizică se poate înţelege fie izolarea fenomenelor de unii factori perturbatori (cazul căderii corpurilor în camere cu vid), fie menţinerea constantă a unor factori, cum ar fi presiunea, temperatura etc. Cercetătorul din ştiinţele naturii are, de regulă, la dispoziţie instrumente şi aparate adecvate pentru realizarea ambelor tipuri de acţiuni.

Pentru a circumscrie mai exact experimentul ca metodă ştiinţifică de cunoaştere, mai menţionez că şi aici, ca şi în cazul observaţiei, acţiunea respectivă trebuie să aibă un caracter eminamente de cunoaştere, în două sensuri: atît prin premisele sale, cît şi prin finalitatea ei. Ambele sublinieri sînt necesare pentru a evidenţia, pe de o parte, faptul că orice experiment ştiinţific porneşte de la o ipoteză (sau mai multe), care derivă dintr-o teorie, mai mult sau mai puţin elaborată şi conştientizată, a fenomenelor cercetate. Pe de altă parte, finalitatea ştiinţifică este o caracteristică ce evidenţiază diferenţa dintre experiment şi tot soiul de "experienţe" care sînt atît de des întîlnite în viaţa socială.

După părerea mea, o metodă ştiinţifică ce se doreşte experiment trebuie să tindă, indiferent de domeniul de aplicare, spre realizarea tuturor condiţiilor amintite mai sus.

De multe ori prin experiment se înţelege numai intervenţia activă, menţionată la punctul (i). Nu e greu de văzut că o asemenea concepţie duce în fapt la identificarea experimentului sociologic cu aproape orice acţiune socială, cu orice activitate practică transformatoare. Confuzia cea mai gravă ce se poate face (şi se face foarte des) în sensul acesta este între experiment şi activitatea de conducere socială. Orice act de conducere înseamnă o modificare a socialului; orice dispoziţie, măsură, reglementare juridică etc. este o intervenţie în desfăşurarea fenomenelor sociale, cu consecinţe mai evidente sau mai ascunse, mai apropiate sau mai îndepărtate, mai previzibile sau mai neaşteptate etc., dar pentru că tuturor acestora le lipseşte , în genere, controlul factorilor şi finalitatea de cunoaştere, ele nu pot fi socotite experimente. De aici nu trebuie să se tragă concluzia că experimentul

n-ar putea sau n-ar trebui să aibă aplicaţii practice. Aceasta este însă o însuşire care nu este esenţială, definitorie pentru actul ştiinţific ca atare. Desigur că ştiinţa se dezvoltă pe seama activităţii practice, fiind influenţată de solicitările acesteia, că ştiinţa are finalităţi foarte practice ş.a.m.d., dar aplicaţiile practice ale ştiinţei nu trebuie confundate cu ştiinţa însăşi.

Alteori se renunţă în definirea experimentului tocmai la intervenţia activă (i). Motivul nu este greu de înţeles şi se bazează pe ideea că cercetătorul poate profita de anumite schimbări care se produc fără intervenţia sa. Un cutremur, o inundaţie sau alte calamităţi naturale pot produce schimbări în viaţa socială, care se cer studiate. Cu toate acestea, nu cred că asemenea situaţii pot fi numite experimente. Mai întîi, pentru că nici în ştiinţele naturii nu se foloseşte acest termen cînd se studiază modificările produse de natura însăşi. Aşa de pildă, nu am sesizat nicicodată că astronomii ar numi eclipsele de soare sau de lună experimente, cu toată înfrigurarea şi interesul cu care sînt aşteptate evenimentele respective. In al doilea rînd, faţă de ştiinţele naturii, unde schimbările sînt în general mai lente, stabilitatea sistemelor mai mare, ştiinţele sociale au, de regulă, de a face cu transformări mult mai evidente şi atunci, practic, orice studiu care urmăreşte un sistem în diacronia sa ar putea fi numit experiment.

Probabil însă că discuţia aceasta ar putea fi cu mult mai folos continuată, pentru că argumentele mele vor fi mai uşor de înţeles, după ce analizăm, chiar dacă foarte schematic, structura situaţiilor experimentale.

2. Schema logică a experimentului

La ora actuală, schemele experimentale au ajuns la complexităţi şi subtilităţi nebănuite de omul de rînd, tocmai cu scopul de a se realiza cît mai fidel cerinţele din definiţie, mai ales cea referitoare la controlul factorilor. Dar oricît de sofisticate ar fi formele de experimentare, ele pornesc de la o schemă relativ simplă, pe care o voi expune în cele ce urmează.

Ne vom plasa în cazul elementar al testării legăturii

dintre două fenomene, din care unul, notat cu y, este considerat a fi efectul sau consecinţa celuilalt, notat cu x. Pentru simplificarea limbajului, vom numi pe x variabilă independentă iar pe y dependentă.

In afara acestor factori, mai trebuie avute în vedere încă două mulţimi de variabile:

- grupul, pe care-l notăm cu A, ce cuprinde toţi acei factori care, pe baza teoriei sau a cunoştinţelor noastre empirice, considerăm că pot avea o influenţă semnificativă asupra lui y sau asupra relaţiei dintre x şi y;

- grupul , pe care-l notăm cu B, ce cuprinde restul factorilor, a căror influenţă o putem considera nulă sau, în orice caz, de un ordin de mărime net inferior celui al factorilor din A.

In continuare, experimentul mai presupune observarea situaţiilor în minimum două momente de timp:

- t_o momentul iniţial, anterior realizării acţiunuii modificatoare;

- t_n, momentul final, după realizarea acţiunii, cînd se evaluează efectul lui x asupra lui y.

Se ajunge la următoarele două scheme:

	t_o	t_n
(I)	A_o, B_o, y_o	x, A_n, B_n, y_n
(II)	x_o, A_o, B_o, y_o	x_n, A_n, B_n, y_n

Primul caz corespunde situaţiei cînd variabila x este efectiv introdusă, factorul respectiv neexistînd la momentul t_o al doilea, cînd variabila x suferă o transformare, trecînd de la x_o la x_n. Pentru ca situaţiile noastre să îndeplinească şi a doua condiţie fundamentală a experimentului, este nevoie ca factorii din complexul A să fie controlaţi, adică acţiunea lor să nu se modifice între t_o şi t_n. Vom nota prescrurtat aceasta prin:

$$A_o = A_n$$

subliniind că semnul egal indică aici o identitate de stare şi nu o egalitate în sens matematic a două mărimi.

Din situaţia (I) va rezulta deci că, dacă toţi factorii A sînt constanţi, schimbarea lui y de la starea y_o la y_n se datorează introducerii factorului x. Vom nota această schimbare cu

$$\Delta_y = y_n - y_o$$

unde semnul minus poate însemna efectiv o diferenţă matematică, dacă factorul y este măsurabil, sau numai o diferenţă calitativă de stări, dacă nu este măsurabil.

In situaţia (II) se observă că variaţia lui x, de la starea x_o la x_n, ce poate fi simbolizată prin:

$$\Delta_x = x_n - x_o$$

produce variaţia lui y:

$$\Delta_y = y_n - y_o.$$

Desigur că în acest caz putem urmări relaţia dintre x şi y nu numai în cele două momente, iniţial şi final, ci şi pe tot intervalul de timp cît durează experimentul. Astfel, va rezulta o relaţie de dependenţă între x şi y, ce poate fi pusă sub formă unui tabel al observaţiilor:

Stările lui x:	x_o	x_1	x_2	. . .	x_n
Stările lui y:	y_o	y_1	y_2	. . .	y_n

dependenţă care, în cazul cînd x şi y sînt măsurabile, poate lua forma unei funcţii matematice:

$$y = f(x)$$

Se mai observă că cele două cazuri, (I) şi (II), corespund la două dintre cunoscutele "canoane" ale lui J.St.Mill. Primul descrie metoda <u>diferenţei</u>, dacă prin fenomenul efect înţelegem tocmai variaţia lui y, iar al doilea, metoda <u>variaţiilor concomitente</u>. Intr-adevăr, formulele (I) şi (II) pot fi scrise:



	t_o	t_n
(I)	A_o, B_o, y_o	$x, A_o, B_n, y_o + \Delta_y$
(II)	x_o, A_o, B_o, y_o	$x_o + \Delta_x, A_o, B_n, y_o + \Delta_y$

Din (I) rezultă că $\underline{x}$ produce pe Δ_y iar din (II) că Δ_x generează pe Δ_y. Tot din schematizarea de mai sus apare ca o concluzie firească faptul că metoda diferenţelor este un caz particular al celei a variaţiilor concomitente, şi anume atunci cînd $x_o = 0$ şi cînd pentru (II) se au în vedere numai momentele t_o şi t_n.

Aşadar, din cele patru canoane propuse de Mill, două se dovedesc adecvate situaţuilor experimentale, rezultînd deci că celelalte, metoda concordanţei şi cea a reziduurilor, sînt specifice observaţiei. De aici nu rezultă însă că cele care se potrivesc experimentului nu pot fi folosite în cazul observaţiei, ci dimpotrivă că ele au virtuţi cognitive superioare în observaţie tocmai pentru că reflectă logica experimentului. Pentru ca metoda diferenţei sau cea a variaţiilor concomitente să se aplice în cazul observaţiei este nevoie de întrunirea a două condiţii:

-factorii $\underline{A}$ să rămînă neschimbaţi;

--observatorul săi aibă, în plus, şansa de a asista la aparaţia neprovocată a fenomenului $\underline{x}$ (pentru cazul I) sau la variaţia acestuia (pentru II)-

Prin urmare, aplicarea în sine a canonului diferenţei sau al variaţiilor concomitente nu înseamnă experiment; se ajunge la experiment numai în măsura în care introducerea sau variaţia factorilor independenţi este provocată, asigurîndu-se în acelaşi timp controlul factorilor $\underline{A}$. Renunţarea la condiţia care cere manipularea variabilei independente are drept consecinţă, prin slăbirea definiţiei, lărgirea sferei noţiunii de experiment pînă la cuprinderea în ea a unor proceduri care nici pe departe nu mai seamănă cu experimentul tradiţional. Iar denumirea unor astfel de proceduri cu termenii de "experimente indirecte" sau "semiexperimente" nu face decît să creeze confuzie, mai cu seamă atunci cînd unii autori le acordă, implicit sau explicit, statutul de experiment.

Un exemplu de asemenea experimentare indirectă este aşa-numita "experimentare ex post facto" . "Ea constă, spune M.Duverger (1964), în cercetarea înlănţuirii factorilor care conduc la o situaţie dată, după ce aceasta s-a produs" (p.367). Voi schiţa acest aşa-zis experiment tocmai pentru a se vedea cît de departe ne aflăm de o situaţie experimentală propriu-zisă. De fapt, există aici două tipuri de abordare : de la cauză spre determinarea efectelor sale şi de la efect spre determinarea cauzei. Pentru ilustrare, voi cita chiar exemplele date de Duverger, provenind din cercetările sociologilor americani.

Primul exemplu urmăreşte stabilirea efectului pe care instrucţia îl are asupra reuşitei sociale. Un cercetător a pornit de la un lot de 1.194 subiecţi, care, cu 9 ani mai înainte, au absolvit o aceeaşi şcoală . Dintre ei o parte şi-au continuat studiile cîţiva ani iar alţii au intrat în cîmpul muncii (variabila independentă). Fiecare subgrup a fost astfel redus încît loturile rămase să fie perfect echilibrate din punctul de vedere al originii sociale, al nivelului intelectual, al naţionalităţii de origine etc. (controlul factorilor). Pînă la urmă s-au ajuns la cîte 145 de indivizi în fiecare grup şi s-a evaluat reuşita socială a indivizilor din cele două loturi (variabila dependentă).

Al doilea exemplu - pentru ilistrarea demersului invers - se referă la o cercetare făcută în Minnesota, avînd ca obiectiv compararea a 5o delinvenţi cu 5o de nondelicvenţi, identici cu primii din punctul de vedere al unităţii şcolare absolvite, al nivelului de inteligenţă, al situaţiei părinţilor etc. S-a încercat urcarea în trecut spre a descoperi cauza delinvenţei.

Asemenea situaţii "experimentale" se pot imagina foarte uşor, dar mie mi se pare că ele nu se deosebesc de simpla observaţie (în sensul în care Durkheim înţelegea metoda comparativă), decît poate printr-o grijă deosebită purtată controlului factorilor. Dar oare se mai practică astăzi observaţii în care controlul factorilor să nu stea în atenţia cercetătorului ? Iar fiindcă sociologia este o disciplină care apelează cel mai des la o abordare statistică a fenomenelor şi proceselor studiate, nu este lipsit de interes de subliniat că o direcţie fundamentală în dezvol_

tarea statisticii în ultimul secol,a fost tocmai aceea care a furnizat instrumente eficiente de control statistic al factorilor E suficient să amintesc aici de analiza corelaţiilor parţiale şi multiple, analiza de regresie, analiza path şi de dependenţă, analiza multivariată a caracteristicilor calitative etc., care toate au ca scop, în ultimă instanţă, să "purifice" relaţia dintre două variabile de acţiunea perturbatoare a altor factori. De fapt orice tabel statistic, care asociază sau corelează simultan cel puţin trei variabile, poate fi interpretat în acelaşi scop.

In concluzie, voi afirma că o cercetare nu are caracter de experiment decît dacă îndeplineşte condiţiile definitorii precizate mai sus. Dacă nu intervine activ în generarea factorului independent, atunci cercetătorul procedează doar la o observaţie, care , chiar dacă e mai riguroasă prin control, tot observaţie rămîne. Dacă nu realizează un control suficient şi nu urmăreşte scopuri ştiinţifice (aceste două elemente fiind, de regulă, legate), atunci activitatea sa se confundă cun un tip sau altul de practică socială.

3. Particularităţile experimentării în sociologie şi în ştiinţele socio-umane, în general

Unii autori încearcă, aşa cum s-a văzut, să găsească deosebiri esenţiale între experimentul din ştiinţele sociale şi cel din ştiinţele naturii; alţii deosebiri mai puţin importante. Din păcate, cînd se vorbeşte despre astfel de diferenţe rareori se precizează pe ce plan se consideră cele două experimente ca fiind diferite: pe cel logico-epistemologic sau pe cel al realizării practice. Altfel spus, trebuie precizat, cînd vorbim de deosebiri, dacă experimentul social prezintă particularităţi specifice în structura sa esenţială, deci ar necesita o definiţie deosebită, sau experimentul fiind unul în orice ştiinţă, atunci cînd ne referim la realitatea socio-umană apar dificultăţi în aplicarea experimentului, de unde pot rezulta eventual anumite forme specifice de realizare, care nu afectează însă cu nimic structura sa generală.

Poziţia mea cred că a reieşit deja suficient de clar din

cele spuse pînă aici: consider că, din punctul de vedere al structurii sale logice şi al trăsăturilor definitorii, experimentul este acelaşi indiferent de domeniul de aplicare. Diferenţieri se pot ivi doar în modalităţile practice de realizare a acestei acţiuni de cunoaştere. Dacă prin definiţie experimentul este ceva în fizică şi altceva în sociologie, atunci s-ar putea ridica - şi se ridică - pe bună dreptate întrebarea de ce asemenea metode trebuie să se mai numească ambele experimente.

Pe de altă parte, se cuvine menţionat că îndeplinirea riguroasă a tuturor condiţiilor (mai ales cea privind controlul factorilor) este mai degrabă un ideal decît o realitate, chiar în ştiinţele naturii cu obiectul cel mai simplu, cum este fizica. Adeseori, ca în multe alte probleme de metodologie, se face greşeala de a se identifica experimentul în ştiinţele naturii cu idealul experimentării, contrapunînd apoi acestuia experimentul social. Adevărul este că şi ştiinţele naturii se deosebesc în funcţie de măsura în care reuşesc să se apropie de acest ideal. Nu toate ştiinţele beneficiază în aceeaşi măsură de posibilitatea provocării unor fenomene, de posibilitatea controlului diferiţilor factori sau a integrării rezultatelor într-o textură teoretică solidă. Rezultă, prin urmare, că deosebirea dintre observaţie şi experiment nu este una absolută. Există, cum spuneam, observaţii care tind să înglobeze cît mai mult cerinţele experimentului, aşa cum există multe experimente imperfecte, din punctul de vedere al realizării lor efective. Aceasta nu înseamnă însă că trebuie să ştergem orice graniţă între observaţie şi experiment. Aceasta nu înseamnă că avem dreptul să decretăm experiment orice observaţie mai sofisticată numai pentru a putea, chipurile, depăşi condiţia unei ştiinţe cu slabe posibilităţi experimentale. De altfel, există ştiinţe foarte venerabile, ca astronomia, care nu apelează deloc la metoda experimentală şi prin aceasta nimeni nu le pune la îndoială nivelul de ştiinţificitate al propoziţiilor.

Cît priveşte sociologia - şi în general ştiinţele sociale - există dificultăţi metodologice generale în abordarea obiectului lor, despre care am vorbit la momentul potrivit. Nu le voi mai repeta aici, ci voi încerca să menţionez forma concretă în care ele se prezintă atunci cînd urmărim realizarea unui experiment Pentru o oarecare sistematizare, o să mă ghidez după elementele

utilizate în definirea acestei metode.

Mai întîi, se impun atenţiei o serie de obstacole legate de intervenţia activă în viaţa socială, de introducerea unor evenimente noi în viaţa unor comunităţi umane. Pentru omul de ştiinţă dificultăţile respective cresc, poate mai mult decît proporţional, pe măsura ridicării de la micro la macrosocial. Asemenea intervenţii întîmpină obstacole, să le spunem, materiale, în sensul că cercetătorul nu are pur şi simplu puterea de a interveni; este cît se poate de evident că, în măsura în care se implică în experiment grupuri cît mai mari de oameni, el trebuie să facă apel la organele conducătoare, politice şi administrative, şi să primească nu numai aprobarea dar şi sprijinul lor efectiv pentru realizarea mutaţiilor dorite. Un alt gen de obstacole îl reprezintă cele morale: una este să experimentezi pe cobai şi alta pe pielea semenilor tăi. Chiar dacă se poate presupune cu mult temei că experimentul nu va influenţa semnificativ viaţa celor investigaţi, elementarele reguli ale eticii ne cer să aducem la cunoştinţa şi să primim consimţămîntul acestora pentru realizarea experimentului.

Dar cele mai serioase greutăţi în calea experimentării în ştiinţele sociale apar la nivelul controlului factorilor. Se pot distinge şi aici mai multe planuri pe care se manifestă aceste dificultăţi.

Se cuvine subliniată, în primul rînd, complexitatea fenomenelor sociale, interdependenţele extrem de numeroase de aceeaşi intensitate sau de intensităţi apropiate, ceea ce nu se întîmplă în aceeaşi măsură în ştiinţele fizice. Aş reaminti celebrul experiment asupra căderii corpurilor în camera cu vid. Atunci cînd "controlăm" densitatea atmosferică, în sensul că eliminăm aerul din cîmpul experienţei, vom controla singurul factor cu adevărat perturbator care influenţează căderea corpurilor; în lipsa acestui factor, legea rămîne şi se manifestă în toată puritatea ei. Poate că mai sînt şi alţi factori care, la o analiză extrem de minuţioasă, ar putea manifesta o influenţă asupra căderii corpurilor, dar influenţa acestora se va dovedi de un ordin de mărime atît de mic încît se poate ca uşurinţă neglija. Să încercăm să găsim asemenea situaţii în societate! Cîţi factori trebuie controlaţi pentru a ajunge la explicarea delicvenţei, a mobilităţii sociale, a

migrației etc. ? Si este evident că, pe măsură ce numărul factorilor ce trebuie controlați este mai mare, cresc și dificultățile respectivului control.

Pe de altă parte, în științele naturii se utilizează aparate, instrumente care permit izolarea fenomenului asupra căruia experimentăm. In felul acesta controlul se poate realiza eficient și, la rîndul său, poate fi controlat, fiind astfel siguri că realmente are loc controlul dorit. Izolarea în științele sociale nu beneficiază, de regulă, de asemenea instrumente valide și ca atare se poate realiza doar în mod cu totul precar, rămînînd de multe ori doar la nivelul intențiilor, fără a putea fi verificată.

In al treilea rînd, este vorba de existența unui factor ce nu se întîlnește în științele naturii, factor care ridică dificultăți absolut inedite în fața cercetătorului social, mai cu seamă în legătură cu controlabilitate lui: mă refer la realitatea că oamenii știu că sînt supuși experimentului, deci că ei sînt subiecți ai investigației. Voi nota acest factor cu <u>C</u>. Acțiunea lui <u>C</u>, în anumite situații experimentale, poate fi atît de importantă încît să estompeze aproape total raportul dintre factorii studiați în cadrul experimentului.

Un exemplu mai mult decît semnificativ în acest sens îl constituie unul dintre celebrele experimente de la Hawthorne, desfășurate de Elton Mayo și echipa sa la uzina Western Electric la sfîrșitul anilor '20. E vorba de experimentul privind influența iluminării locului de muncă asupra randamentului (De fapt acest experiment a început înainte de sosirea lui E.Mayo la Hawthorne, dar el este inclus, de regulă, în șirul de experimente desfășurate în cadrul programului respectiv). S-au luat două grupuri de muncitori, unul la care iluminatul a rămas constant și altul la care a variat. La ambele echipe productivitatea muncii a crescut simțitor; la cea cu iluminare variabilă, productivitatea a sporit atît la mărirea gradului de iluminare cît și la diminuarea acestuia pînă la limita vizibilității. Este deci evident că factorul care a dus la creșterea productivității a fost însăși conștiința indivizilor că sînt observați, că se găsesc într-o situație cu totul specială. Deci relația de dependență presupusă între <u>y</u> (productivitate) și <u>x</u> (iluminat) a rămas finalmente ascunsă datorită factorului <u>C</u>.

S-ar putea desigur înmulţi la nesfîrşit exemplele care să confirme aprecierea că dificultăţile controlului sînt foarte accentuate în experimentele din ştiinţele socio-umane, dar consider că lucrurile sînt deja suficient de clare. Se ridică atunci firesc întrebarea dacă această cerinţă fundamentală a experimentului - controlul - este realizabilă(sau realizabilă într-o măsură satisfăcătoare) pentru a putea vorbi de experimente şi în acest cîmp al cunoaşterii. Răspunsul este afirmativ fiindcă, din fericire, inventivitatea omului este fără limite şi, în cazul de faţă, s-a găsit o procedură satisfăcătoare de control fără a se acţiona direct asupra factorilor controlaţi. Este vorba despre experimentul cu grup (grupuri) de control, în care controlul se realizează în mod indirect. Ar fi de precizat că această tehnică experimentală nu este specifică ştiinţelor sociale ; ea se foloseşte în multe sectoare ale ştiinţei, peste tot unde controlabilitatea este redusă. Biologia, ştiinţele agricole şi cele medicale sînt numai cîteva exemple de asemenea discipline pentru care experimentul cu grup de control este o metodă curent întrebuinţată.

Schema logică a experimentului de tip (I), cu grup de control este:

	t_o	t_n
G:	A_o, B_o, y_o	x_n, A_n, B_n, y_n
G':	A_o, B'_o, y_o	A_n, B'_n, y'_n

unde prin G am notat grupul experimental iar prin G' grupul de control.

Grupul de control G', se alege de aşa manieră încît să coincidă în cea mai mare măsură cu grupul experimental, ceea ce înseamnă aceeaşi stare a variabilelor cuprinse în A şi aceeaşi valoare a lui y la momentul iniţial. De asemenea, se mai face supoziţia că, între t_o şi t_n setul factorilor A urmează aceeaşi variaţie în ambele grupuri, trecînd de la starea A_o la A_n.

Se vede deci că, astfel ales grupul de control, într-o asemenea situaţia experimentală, nu mai este nevoie să se pună condiţia:

$$A_n = A_o$$

care, aşa cum am văzut, creează cele mai mari obstacole în calea experimentului social. Intr-adevăr, chiar dacă factorii A variază între cele două momente de timp, din ipotezele menţionate mai sus rezultă că efectul introducerii factorului x numai în grupul experimental este marcat în diferenţa de stări la care ajunge factorul dependent, y în cele două grupuri la momentul final. Deci, formal, efectul lui x va fi:

$$y_n - y_n'$$

şi nu

$$y_n - y_o$$

ca în cazul experimentului fără grup de control. In adevăr, punînd

$$y_n - y_o = y_n - y_n' + y_n' - y_o$$

se vede că schimbarea variabilei în grup experimental $(y_n - y_o)$ se descompune în suma dintre variaţia lui y în grupul de control $(y_n' - y_o)$ şi diferenţa valorilor rezultate în cele două grupuri $(y_n - y_n')$. Admiţînd că variaţia în grupul de control se datorează trecerii factorilor A din starea A_o în A_n şi că aceeaşi modificare se produce şi în grupul experimental, rezultă că efectul introducerii noului factor x este $y_n - y_n'$.

Un exemplu simplu de astfel de experiment poate fi luat din domeniul pedagogiei. Să presupunem că un profesor doreşte să evalueze eficienţa unei metode noi de instrucţie. Va alege două clase de elevi cît mai apropiate din punctul de vedere al structurii şi cu rezultate identice la disciplina de studiu respectivă. La o clasă va utiliza nou tehnologie pedagogică iar la cealaltă va rămîne la vechea metodă folosită. Dacă cele două colectivităţi au fost corect alese şi pe parcursul experimentului (de pildă un an şcolar) nici un grup nu a fost supus influenţei vreunui alt factor specific, atunci diferenţa dintre rezultatele şcolare ale celor două clase poate fi pusă pe seama factorului nou introdus.

Pentru experimentul de tip (II), schema formală, în cazul cînd se foloseşte un grup de control, va fi:

$$
\begin{array}{llll}
 & t_o & & t_n \\
G: & x_o, A_o, B_o, y_o & & x_n, A_n, B_n, y_n \\
G': & x_o, A_o, B'_o, y_o & & x_o, A_n, B'_n, y'_n
\end{array}
$$

Diferența dintre cele două grupuri este aici marcată de faptul că în cel de control factorul $\underline{x}$ rămîne constant, pe cînd în cel experimental el este variabil. Efectul variației lui $\underline{x}$ va fi și în acest caz diferența dintre stările finale ale lui $\underline{y}$ în cele două grupuri.

Se înțelege că formele prezentate mai sus reprezintă schemele cele mai simple de experiment, care surprind esența procedurii științifice analizate aici. Ele se pot "complica" în funcție de diferite elemente concrete care intervin în aplicarea lor în practică. De exemplu, în cazul (I), dacă nu se poate asigura identitatea stărilor lui $\underline{y}$ la momentul inițial în cele două grupuri, deci dacă vom avea în $\underline{G'}$ nu y_o ci y'_o, atunci efectul introducerii lui $\underline{x}$ va fi:

$$(y_n - y_o) - (y'_n - y'_o)$$

De asemenea, dacă pentru cazul (II) nu avem posibilitatea să variem efectiv variabila $\underline{x}$ în cadrul <u>unui</u> grup experimental, se pot alege <u>mai multe grupuri experimentale</u>, fiecare fiind caracterizat de o anumită stare a lui $\underline{x}$.

In cazul experimentelor cu grup de control, principala dificultate ce trebuie învinsă este alegerea grupurilor de așa manieră încît să avem un grad suficient de ridicat de certitudine că, exceptînd acțiunea lui $\underline{x}$ în grupul experimental, toate celelalte transformări se vor realiza identic în ambele grupuri. De aceea se urmărește constituirea grupului de control, astfel încît să reflecte cît mai fidel structura grupului experimental, să se ajungă la o cît mai perfectă similitudine a lor. Există, practic, două procedee principale pentru atingerea acestui deziderat:

- procedeul potrivirii pe perechi și
- procedeul probabilistic

In primul caz, pentru fiecare individ din grupul experimental se alege un individ avînd caracteristici identice, în

×

grupul de control (exemplu: acelaşi sex, vîrstă, pregătire şcolară etc.) In al doilea , grupurile se aleg ca nişte eşantioane aleatoare din aceeaşi populaţie, mizîndu-se pe faptul că întîmplarea va potrivi suficient de bine lucrurile, astfel încît să se cuprindă în aceleaşi proporţii anumite tipuri de indivizi în fiecare grup.

Se observă că riscurile sînt mari în ambele situaţii. In catul potrivirii pe perechi, putem controla doar un număr relativ restrîns de factori (aceia care efectiv sînt manifeşti) şi deci putem omite unii foarte importanţi. In al doilea , aşa cum am sugerat deja, ne încredem într-o şansă care ar putea să ne joace cele mai neplăcute feste. Iar probabilitatea să se întîmple aşa este cu atît mai mare cu cît grupurile sînt mai reduse ca dimensiune. Or, în cele mai frecvente cazuri, nu avem posibilitatea de a experimenta decît tocmai cu asemenea grupuri restrînse.

N-am pomenit pînă aici nimic despre factorul C. Este evident că dacă acţiunea sa este considerată a fi semnificativă, e nevoie ca această acţiune să se manifeste cu egală intensitate în grupul experimental şi în cel de control. Adică e bine, dacă cei din G ştiu să participă la un experiment, s-o ştie şi cei din G' şi să reacţioneze la fel. Este tocmai ideea subliniată de A. Mihu (1973), care sugerează administrarea "unui pseudo (sau surogat) x şi în grupul de control" (p.105)

Mai menţionez, în sfîrşit, şi dificultăţile legate de cea de a treia condiţie pe care am subliniat-o pentru experiment: caracterul ştiinţific. Foarte multe obstacole sînt ridicate de gradul scăzut de dezvoltare a teoriei sociologice, ceea ce are repercusiuni asupra alegerii fenomenului de studiat, asupra consistenţei ipotezelor etc. Alte dificultăţi sînt legate de scopul cercetării. Experimentarea în societate presupune aproape întotdeauna o dependenţă a cercetărilor de factorii de decizie, deci subordonarea interesului ştiinţific celui practic, cu toate consicinţele ce pot uneori decurge de aici asupra rigorii ştiinţifice.

4. Experimentul de laborator şi experimentul de teren

Literatura metodologică a ştiinţelor sociale face, de regulă, o distincţie între aceste două tipuri de experiment. In ceea ce mă priveşte, ţin s-o precizez de la bun început, nu mi se pare că o atare diferenţiere s-ar cere cu atîta putere subliniată, datorită faptului că în practică delimitarea nu apare totodeauna prea netă. Dificultăţile în distingerea celor două tipuri este legată de alegerea unui criteriu ferm de departajare. De regulă, este invocat unul sau altul dintre următorii factori, ţinînd cont de o dihotomizare a stării fiecăruia:

(i) Mediul: artificial sau natural
(ii) Izolarea: puternică sau slabă
(iii) Situaţia: inedită sau naturală
(iv) Factorul C: prezent sau absent

Se vorbeşte deci de experiment de laborator, în contrast cu cel natural, atunci cînd subiecţii sînt: urmăriţi într-un mediu artificial şi/sau puternic izolaţi de influenţe exterioare şi/sau supuşi unor situaţii inedite şi/sau conştienţi de faptul că sînt supuşi experimentului.

Este clar că aceste dihotomii sînt prea rigide pentru a se putea identifica întocmai în practică. Faptul că, de exemplu, în cadrul unei întreprinderi, luăm un grup de indivizi şi-l plasăm într-o începere separată înseamnă cu siguranţă alterarea mediului lor natural de muncă, dar nu este vorba nici de condiţii cu totul artificiale. Cît priveşte ultima cerinţă, astăzi se recomandă, din mai multe motive, dar mai ales de ordin etic, ca orice experiment să se facă numai cu acordul subiecţilor, deci, practic, ultimul criteriu este din ce în ce mai puţin operant. Probabil că cea mai evidentă situaţie în care se poate vorbi de experiment de laborator este cea indicată de factorul (iii), adică atunci cînd subiecţii sînt supuşi să joace anumite roluri pe care ei nu le îndeplinesc în viaţa de zi cu zi, căci aceasta implică în fapt şi celelalte cerinţe.

Chiar dacă se acceptă distincţia dintre cele două genuri

de experiment, trebuie precizat că atunci cînd se vorbeşte de experimentul de laborator nu e neapărat nevoie să ne imaginăm că indivizii sînt plasaţi într-o sală pe a cărei uşă scrie "Laborator" că sînt înconjuraţi de tot felul de aparate etc., deşi există evident şi astfel de laboratoare. Ele sînt descrise pe larg în literatura de specialitate - a se vedea, de exemplu, S.Chelcea (1982) - şi nu voi insista asupra modului lor de dotare şi organizare.

Ceea ce aş vrea să subliniez aici este o problemă epistemologică importantă, legată de experiment şi exprimată adesea în discuţiile despre valoarea şi oportunitatea experimentelor de laborator. Se reproşează frecvent acestei din urmă forme de experimentare faptul că ar scoate indivizii din situaţia naturală, plasîndu-i în una artificială, ceea ce ar altera raporturile dintre ei şi s-ar ajunge la concluzii fără nici o aplicabilitate. Iată ce spune în acest sens T.Nicola (1973), care, după ce includea printre elementele ce constituie "esenţa experimentului ştiinţific" starea pură şi crearea artificială de condiţii (p.7), face o cotitură de 18o de grade, susţinînd că experimentul de laborator "este limitat prin capacitatea sa de a surprinde acea autenticitate a realului ...., a sesiza şi desprinde esenţa fenomenului social şi, prin urmare, este limitat în posibilităţile de a generaliza rezultatele dobîndite, de a fundamenta şi construi noi teorii" (p.23). Aş mai aminti că, în aceeaşi lucrare, autorul nega posibilitatea observaţiei de a junge la esenţa fenoemenelor, punînd superioritatea metodei experimentale tocmai pe asemenea studii în stare "pură" (p.4).

Avem de a face în toate criticile de genul celei de mai sus (chiar dacă sînt perfect consecvente) cu o neînţelegere a ceea ce înseamnă o legitate a naturii (socialului). Legea, ca element ce ţine de esenţă - o ştim cel puţin de la Hegel încoace - nu se manifestă în mod vizibil constant în fenomene, acestea trebuind decorticate pentru ajunge la esenţă, Nici pentru fizică nu se întîlnesc în natură condiţiile de laborator, dar pentru asta nimănui nu i-a trecut prin cap să critice experimentul de laborator pentru că nu reproduce condiţiile reale, nimeni nu respinge, de exemplu, legea căderii corpurilor în vid pentru motivul că pe pămînt nu există vid şi deci nicăieri fulgul nu va ajunge la sol odată cu piatra, cînd le dăm drumul să cadă în acelaşi moment şi

de la aceeaşi înălţime. A critica experimentul că nu surprinde "autenticitatea realului" înseamnă a nu sesiza esenţa acestei metode de cercetare ştiinţifică.

Leon Festinger(1963) argumentează şi el în acelaşi sens. Experimentul de laborator "nu are şi nu poate avea pretenţia de a reproduce o situaţie a vieţii reale. Dacă vrem să studiem un aspect al unei situaţii realmente trăite, ar fi o nebunie să vrem să aranjăm într-un raport condiţii copiate de pe această situaţie. De ce să nu se meargă direct la situaţie şi să se studieze aşa cum este ea însăşi. Experimentul de laborator vizează crearea unei situaţii în care operarea cu anumite variabile va putea fi pusă în lumină datorită unor condiţii speciale, bine identificate şi definite în mod clar. Puţin importantă că o astfel de situaţie nu se întîlneşte în viaţa reală" (p.166).

Dacă ar fi să vorbim despre limitele aşa-ziselor experimente de laborator, atunci ar trebui să subliniem că eventuala lor irelevanţă se datorează tocmai dificultăţilor de a întruni în mod riguros condiţiile experimentării şi, în speţă, cele referitoare la izolarea şi controlul factorilor. De asemenea, este evident că doar o parte din fenomenele sociale se pretează la astfel de experimentări, întrucît este greu de văzut cum pot fi îndeplinite cerinţele experimentului în cazul unor fenomene ce se manifestă numai prin antrenarea unor mase mari de oameni.

De aici rezultă şi concluzia generală privind posibilitatea, limitele şi oportunitatea experimentării în sociologie. Dacă se vor inventaria cazurile de experimente indicate în literatura ştiinţelor sociale (ceea ce n-am mai făcut aici din lipsă de spaţiu şi pentru a nu repeta lucruri pe care oricine le poate citi în lucrările indicate la bibliografie) se va ajunge la următoarele constatări:

a) In cele mai numeroase cazuri, studiile experimentale respective aparţin unor domenii învecinate sociologiei, ca psihologia socială (mai cu seamă problematicii psihosociologice a grupurilor mici), ergonomia, ştiinţa conducerii etc.

b) Aşa-numitele experimente sociale de mare anvergură sînt, aproape în totalitatea cazurilor, de fapt pseudoexperimente, deoarece nu întrunesc în chip satisfăcător condiţiile de bază ale a-

cestei metode ştiinţifice.

De aici nu rezultă o respingere de plano a experimentului ca metodă de cunoaştere sociologică, ci doar sublinierea dificultăţilor imense ce stau în faţa sociologului care ar intenţiona să nu uzeze de acest gen de investigaţie. De aici rezultă, de asemenea, că trebuie să punem în întregime relativa întîrziere în dezvoltarea sociologiei pe seama dificultăţilor în realizarea practică a metodei experimentale. Există, aşa cum am mai spus, ştiinţe ale naturii care nu utilizează experimentul, ceea ce nu le-a împiedicat să aibă o istorie multimilenară şi să ajungă la rezultate remarcabile.

Probabil că sociologii şi alţi specialişti din ştiinţele sociale consideră că atribuirea denumirii de experiment unor serii de tehnici de cercetare care nu au nimic de a face cu metode experimentală ar avea darul să ridice prestigiul investigaţiilor lor. Este aceasta însă o tactică ce, pînă la urmă, se poate întoarce împotriva celui care o utilizează şi a sociologiei în ansamblul său.

PARTEA A III-A

ANALIZA STATISTICA A DATELOR

Cap.VII. CARACTERISTICI

Noţiunea de "caracteristică", încetăţenită de multă vreme în limbajul statisticii descriptive, nu reprezintă altceva decît trăsătura, însuşirea etc. în funcţie de care s-a realizat o clasificare sau o măsurare propriu-zisă în cadrul unei mulţimi de indivizi ce constituie eşantionul sau populaţia investigată de cercetător.

Pe baza unor consideraţii de ordin foarte practic - în sensul că ele provin exclusiv din modalităţile tehnice de lucru cu datele -, în statistică se face o distincţie netă între caracteristicile calitative şi cele cantitative, primele corespunzînd acelor trăsături care s-au abordat cu ajutorul scalelor nominală şi ordinală, iar ultimele celor ce corespund scalelor de intervale şi de rapoarte. In esenţă, deosebirea aceasta se datorează imposibilităţii sau posibilităţii de a utiliza valorile numerice, atribuite treptelor scalei, în operaţii matematice obişnuite. Chiar dacă, aşa cum am arătat, există diferenţe mari între scalele nominale şi cele ordinale sau între cele de intervale şi de rapoarte, chiar dacă, pe de altă parte, scala ordinală presupune deja manifestarea cantitativă a trăsăturii studiate, apropiindu-se deci de cea de intervale, în operaţiile statistice distincţia menţionată între cele două tipuri de caracteristici este foarte tranşantă, traducîndu-se - cum vom vedea imediat - în procedee diferite de tratare a datelor.

In scopul simplificării limbajului, voi folosi, în cele ce urmează, pentru caracteristicile cantitative termenul, de asemenea încetăţenit, de variabilă statistică , sau, pur şi simplu, de variabilă.

1. Distribuţii de frecvenţă

Aplicarea unei scale nominale sau ordinale la o colectivitate statistică ne conduce imediat la o clasificare a indivizilor, adică la o repartizare a acestora pe treptele scalei respective. Spre deosebire de operaţiile de clasificare folosite în general în diferite ştiinţe, clasificările de tip statistic implică şi precizarea cuantumului de indivizi ce se plasează pe diferite trepte ale scalei. Prin urmare, noţiunea de caracteristică implică, aşa cum am subliniat de la început, nu numai o trăsătură sau, eventual, un pachet de trăsături după care se efectuează clasificarea ci şi modalitatea în care indivizii se distribuie în clasele respective. Tocmai acest aspect cantitativ, legat de numărarea indivizilor din diferite clase şi de operaţiile cu numerele astfel obţinute, constituie conţinutul tuturor procedurilor statistice referitoare la caracteristicile calitative.

In modul cel mai general vorbind, construcţia unei caracteristici calitative, A, înseamnă realizarea unei corespondenţe între o mulţime:

$$\mathcal{A} = \{A_1, A_2, \ldots, A_s\}$$

constituită din clasele $A_1, A_2, \ldots, A_s$ ale scalei şi o mulţime:

$$\mathcal{K} = \{k_1, k_2, \ldots, k_s\}$$

de numere naturale, ale cărei elemente arată cîţi dintre indivizii populaţiei studiate se plasează pe fiecare treptă a scalei. Mulţimea $\mathcal{K}$, ale cărei elemente se numesc frecvenţe sau frecvenţe absolute, reprezintă distribuţia de frecvenţă a caracteristicii A.

O asemenea corespondenţă se scrie, de regulă, sub forma unui tabel ca cel de mai jos:

Clasele	A_1	A_2 ...	A_s	Total
Frecvenţele	k_1	k_2 ...	k_s	n

Pe prima linie se trec simbolurile claselor constituite iar pe a doua frecvenţele , k_i , a căror sumă trebuie să fie egală cu n, volumul populaţiei studiate:

$$k_1 + k_2 + ... + k_s = n$$

sau, prescurtat

$$\sum_{i=1}^{s} k_i = n$$

Relaţia de mai sus provine evident din principiul clasificării, conform căruia fiecare individ din populaţie trebuie să se plaseze în una şi numai în una din clase.

Forma cea mai simplă a unei caracteristici calitative este cea dihotomică, realizată atunci cînd populaţia se divide numai în două clase. De obicei, o astfel de caracteristică poartă numele de atributivă sau de atribut, deoarece aproape întotdeauna se poate considera că indivizii plasaţi în una din clase posedă un atribut, pe cînd ceilalţi nu-l posedă. In acest caz, se poate folosi pentru clase notaţia A (pentru cei care au calitatea respectivă) şi $\bar{A}$ (pentru cei care n-o au). Tabelul devine atunci:

Classele	$A(A_1)$	$\bar{A}(A_2)$	Total
Frecvenţele	k_1	k_2	n

Pentru variabilele statistice situaţia este puţin diferită. Elementele scalei sînt numere care se obţin printr-o operaţie de măsurare propriu-zisă sau prin alte proceduri de atribuire. Aceste valori ale scalei pot fi în număr finit sau infinit , plasate pe un interval ce merge de la valoarea cea mai mică la cea mai mare. Cînd o variabilă acoperă, potenţial vorbind, cu valorile sale un întreg interval [a,b], ea se numeşte continuă , iar atunci cînd ia numai anumite valori, de regulă în număr finit, discretă.

Este evident că în acţiunile concrete de cercetare nu vom întîlni decît variabile discrete, deoarece mijloacele noastre de investigaţie şi percepţia umană nu pot surprinde o infinitate (cor-

tinuă) de nuanţe. Aceasta nu înseamnă că trebuie să renunţăm la distincţia dintre cele două tipuri de variabile. De pildă, vîrsta, măsurată pe axa timpului, este clar o variabilă continuă, ce poate lua valori de la 0 la loo şi ceva de ani, deşi practic este imposibil să precizăm vîrsta cuiva cu o exactitate de ordinul secundelor. Dimpotrivă, dimensiunea unei gospodării este o variabilă discretă clasică, singurele valori posibile fiind: 1,2,3 etc. Chiar şi unele variabile discrete, prin definiţie, dar cu un număr foarte mare de valori pot fi asimilate celor continue. De exemplu, retribuţia, presupunînd că nu se calculează în subdiviziuni monetare mai mici de 1 leu, atinge o mulţime extrem de mare de valori, de ordinul miilor, şi ca atare poate fi practic considerată o variabilă continuă, dacă interesele studiului cer acest lucru.

In cazul variabilelor discrete cu un număr relativ mic de valori, acestea din urmă (valorile) se substituie, formal, simbolurilor claselor şi situaţia poate fi prezentată ca şi pentru caracteristicile calitative:

Valori	x_1	x_2	...	x_s	Total
Frecvenţe	k_1	k_2	...	k_s	n

Dacă variabila este continuă sau cu un număr foarte mare de valori, sistematizarea datelor devine practic imposibilă dacă în tabel s-ar trece valoarea exactă a fiecărui individ în parte. In astfel de cazuri, se apelează la un artificiu, şi anume se construiesc clase prin gruparea tuturor valorilor dintr-un subinterval; se obţin astfel *caracteristici cu intervale de grupare*. Este de notat că aceste caracteristici îşi pierd într-o oarecare măsură atributul de *variabile*, întrucît fiecare interval-clasă nu mai este caracterizat de *o* valoare numerică aşa cum cerea definiţia, ci de o mulţime de valori. Intervalele corespund mai degrabă unor *calităţi*, avînd în vedere că toţi indivizii plasaţi într-un acelaşi interval sînt consideraţi *identici*, din punctul de vedere al caracteristicii studiate. Dar cum trecerea de la variabile la caracteristici calitative înseamnă de fapt un regres (pentru că primele se pretează la infinit mai multe operaţii şi analize statistice), atunci cînd cercetarea nu se opreşte la simpla descriere

a caracteristicilor, se procedează la o revenire spre cantitativ, atribuindu-i-se fiecărui interval o singură valoarea numerică, de regulă una centrală intervalului.

Prin urmare, trecerea de la variabile propriu-zise la caracteristici cu interval de grupare şi revenirea la variabile presupun două operaţii care se expun unui grad mai mare sau mai mic, dar totdeauna nonneglijabil, de convenţionalitate: alegerea limitelor intervalelor şi a valorilor caracteristice pentru acestea. La prima operaţie intervin o serie de elemente de apreciere ce pot deriva din teoria asupra fenomeneului analizat şi/sau din distribuţia empirică a indivizilor după variabila originară. Astfel, clasele reprezentate de intervale pot fi impuse de considerente ţinînd de identitatea calitativă a indivizilor, care, chiar dacă iau valori diferite, pot fi consideraţi de aceeaşi natură. De pildă, dacă variabila este "vîrsta" şi o utilizăm într-o cercetare ce vizează aspecte legate de activităţile şcolare, putem forma intervale de vîrstă care să corespundă preşcolarilor, elevilor din învăţămîntul primar, gimnazial, şi liceal, studenţilor etc. Dacă nu avem nici un motiv de asemenea natură care să sugereze grupările, atunci vom folosi asemenea intervale (nu neapărat de lungime egală) încît să se cuprindă în fiecare un număr suficient de mare de cazuri pentru a putea continua analizele.

Există şi unele formule - empiric stabilite - care ne pot ajuta în evaluarea lungimii optime a intervalelor, cum este cea a lui Sturges:

$$i = \frac{x_{max} - x_{min}}{1 + 3,322.\lg n}$$

unde i este lungimea intervalului, x_{max} şi x_{min}, cea mai mare şi respectiv cea mai mică valoare luată în populaţie iar lgn logaritmul zecimal din numărul total de indivizi. Valoarea unor asemenea formule este destul de limitată; uneori ele pot fi folosite cu folos, alteori nu. De exemplu, dacă variabila x este vîrsta, cu valori între 0 şi 100 ani, pentru un eşantion de 1ooo persoane (lg 1ooo = 3) vom avea:

$$i = \frac{100}{1 + 3,322 \cdot 3} \approx \frac{100}{11} \approx 9$$

Pentru comoditate, se poate lua intervalul de 10 ani şi atunci se obţine o clasificare acceptabilă. In schimb, dacă $\underline{x}$ este retribuţia, cu valori între, să zicem, 1700 şi 5000 lei, pentru acelaşi eşantion de 1000 persoane obţinem $i = 3300 : 11 = 100$ lei, deci 33 de intervale, ceea ce este foarte mult pentru o analiză obişnuită.

Formula lui Sturges presupune că intervalele de grupare sînt egale, lucrul normal pentru că tocmai astfel de situaţii sînt cel mai adesea întîlnite. Dar nu se întîmplă mereu aşa. Uneori indivizii sînt concentraţi foarte puternic într-o anumită zonă a scalei şi se găsesc în număr mic pe alte părţi întinse ale ei. Se cer formate atunci intervale de lungimi inegale, împărţind în sectoare mai mici zona populată şi mai mari cea slab populată. In fine, o ultimă observaţie referitoare la intervale priveşte grija pe care trebuie să o avem atunci cînd fixăm capetele acestora, pentru a putea plasa exact indivizii care iau valori în zona de întîlnire a două intervale. Singurele excepţii de la această regulă o constituie uneori primul şi ultimul interval, care pot rămîne deschise la stînga şi, respectiv, la dreapta, dacă nu se poate preciza valoarea minimă şi maximă.

Revenind la distribuţiile de frecvenţă, trebuie menţionat că cele mai elementare calcule ce se pot face cu ele - indiferent de natura caracteristicii - constau în determinarea <u>frecvenţelor relative</u>. Frecvenţa relativă a unei clase a caracteristicii este fracţiunea (ponderea) indivizilor din populaţia totală care aparţin acestei clase. Dacă notăm cu p_i frecvenţele relative, atunci:

$$p_i = \frac{k_i}{n} \quad (i = 1,2,\ldots,s)$$

adică frecvenţa relativă se obţine prin împărţirea frecvenţei absolute la numărul total de indivizi din populaţie.

Rezultă imediat două proprietăţi importante ale frecvenţelor relative:

a) ele sînt valori pozitive subunitare ($0 < p < 1$);

b) suma lor este egală cu unitatea: $p_1 + p_2 + ... + p_s = 1$

Tot o speţă de frecvenţa relative sînt şi procentele. In loc de a se reduce totalul la unitate, ca în cazul precedent, pentru procente el se reduce la 100. In consecinţă, procentul de indivizi dintr-o clasă se va calcula:

$$\text{procentul} = \frac{k_i}{n} \cdot 100 = p_i \cdot 100$$

Frecvenţele relative şi procentele se utilizează pentru două scopuri fundamentale:

a) Să sesizăm mai uşor care este ponderea diferitelor clase în populaţie. Astfel, dacă într-o clasă avem 348 indivizi, dintr-o populaţie de 4350 , prin calcularea procentului (8%) vom înţelege mai uşor gradul de cuprindere al acestei clase.

b) Mai folositoare sînt frecvenţele relative şi procentele atunci cînd dorim să comparăm caracteristici identice construite asupra unor populaţii de mărimi diferite. Comparaţia este practic imposibilă cu frecvenţe absolute deoarece aceste mărimi au semnificaţii diferite în funcţie de total. Calculul frecvenţelor relative şi al procentelor constituie deci o procedură de standardizare a informaţiei dintr-un tabel statistic. Pe această cale se poate constata că cei 348 de indivizi de mai înainte reprezintă aceeaşi fracţiune, în populaţia de 4350, ca şi alţi 64 indivizi dintr-o populaţie de 800.

Frecvenţele relative (la unitate) se folosesc mai ales ca mărimi intermediare în alte operaţii de analiză statistică la care sînt supuse caracteristicile sau ca valori interpretabile în termeni de probabilităţi. Procentele îşi găsesc cea mai curentă utilizare în prezentarea datelor în raponrte de cercetare, articole, cărţi etc., deoarece ele sînt mai uşor de înţeles de către publicul larg. In adevăr, pentru cei mai mulţi oameni este mai uşor să înţeleagă semnificaţia aserţiunii "în anul cutare, în ţara cutare, în medie din 100 de copii născuţi, 51 au fost băieţi şi 49 fete (sau 51% dintre născuţi au fost băieţi şi 49% fete)" decît a propoziţiei: "în anul cutare, în ţara cutare, ponderea băieţilor în rîndul nou-născuţilor a fost de 0,51 iar a fetelor de 0,49", deşi

informaţia transmisă sub cele două învelişuri formale este absolut identică.

In continuare, voi trece la prezentarea unor exemple concrete, în scopul ilustrării celor spuse pînă aici. Fiind, vorba, în esenţă, de o ilustrare prin intermediul tabeleleor, îmi permit pentru a nu mai reveni cu altă ocazie, să precizez acum cîteva reguli elementare cu privire la utilizarea şi prezentarea tabelelor în cadrul lucrărilor ştiinţifice:

- Fiecare tabel primeşte un număr. Numerotarea poate merge în continuare de la începutul pînă la sfîrşitul lucrării sau poate fi reluată de la 1 în cadrul fiecărui capitol sau secţiuni. In acest din urmă caz, este recomandabil ca, pe lîngă numărul de ordine al tabelului, să se indice şi cel al capitolului, pentru a se facilita trimiterile în text la tabele. De exemplu, scriind "Tabelul 2.7", se va înţelege că este vorba de tabelul 7 din capitolul 2.
- Fiecare tabel are un titlu în care se precizează caracteristica descrisă, populaţia asupra căreia a fost calculată, tipul de frecvenţe sau alţi indicatori folosiţi în tabel. Titlul tabelului urmează după numărul de ordine al tabelului, aceste informaţii putînd fi scrise deasupra sau dedesubtul său, mai aproape de tabel decît de restul textului, pentru a se putea citi uşor.
- Dacă tabelul conţine date ce nu sînt culese de către autor, se va specifica sursa de provenienţă a acestora.
- Capetele liniilor şi coloanele tabelului este bine să conţină o informaţie care, coroborată cu cea din titlu, să permită cititorului să înţeleagă tabelul fără a face apel la textul lucrării.

Să începem cu caracteristicile calitative şi mai exact cu cazul dihotomic. Tabelele 1 şi 2 reprezintă repartiţia populaţiei R.S.R., în 1982, după caracteristicile "sex" şi "mediu", iar tabelul 3 răspunsurile - de forma Da-Nu- la o întrebare de opinie dintr-un chestionar.

Sexul	Bărbaţi	Femei	Total
Frecvenţe	11.o91.465	11.386.238	22.477.7o3
%	49,3	5o,7	loo

Tabelul 1. Populaţia R.S.R. pe sexe în 1982 Sursa: Anuarul Statistic al R.S-R. 1983

Mediul	Urban	Rural	Total
Frecvenţe	11.575.632	lo.9o2.o71	22.477.7o3
%	51,5	48,5	loo

Tabelul 2. Populaţia R.S.R. pe medii în 1982. Sursa: Anuarul statistic al R.S.R. 1983

Răspunsul	DA	NU	Total
Frecvenţe	9oo	3oo	12oo
%	75,o	25,o	loo

Tabelul 3. Răspunsurile la o întrebare din chestionar (date fictive)

Pentru caracteristicile calitative cu mai multe clase, avem, de regulă, două mari situaţii: cînd scala este nominală şi atunci clasificarea apare ca o structurare orizontală a populaţiei şi cînd scala e ordinală, ceea ce implică o structurare pe verticală (sau pe nivele ierarhice) a acesteia. Un exemplu din prima categorie este dat în tabelul 4, iar a doua în tabelul 5.

Starea civilă	Necăsătorit	Văduv	Divorţat	Total
Frecvenţe	149.1o3	5.261	2o.o83	174.447
%	85,5	3,o	11,5	1oo

Tabelul 4. Populaţia masculină din R.S.R. care s-a căsătorit în 1982, după starea civilă înainte de căsătorie. Sursa: Anuarul Statistic al R.S.R. 1983

Răspunsuri	Foarte puţin	Puţin	Moderat	Mult	Foarte mult	Total
Frecvenţe	1oo	4oo	5oo	4oo	2oo	16oo
%	6,25	25,oo	31,25	25,oo	12,5o	1oo

Tabelul 5. Distrubţia răspunsurilor la o întrebare din chestionar. (date fictive)

Următoarele tabele prezintă trei cazuri de caracteristici cantitative. In tabelul 6 apare o caracteristică discretă pentru care sînt indicate exact toate valorile. Tabelele 7 şi 8 cuprind două exemple de caracteristici cu intervale de grupare, ultimul prezentînd şi particularitatea că intervalele de la margine sînt deschise.

Nr.de persoane în gospodărie	1	2	3	4	5	6	7	8	9	Total
Nr.de gospodării	3o	4o	6o	5o	3o	2o	1o	5	5	25o
%	12,o	16,o	24,o	20,o	12,o	8,o	4,o	2,o	2,o	1oo

Tabelul 6. Distribuţia gospodăriilor unei localităţi rurale după numărul membrilor ce le compun(date fictive)



Medii	2 - 2,99	3 - 3,99	4 - 4,99	5 - 5,99	6 - 6,99	7 - 7,99	8- 8,99	9- 1o	Tot.
Nr. de candidați	3o	4o	13o	15o	1oo	75	5o	25	6oo
%	5,6	6,7	21,7	25,o	16,7	12,5	8,3	4,2	1oo,1

Tabelul 7. Distribuția mediilor obținute de candidații la admitere la facultate (date fictive)

Vîrsta	Sub 15 ani	15-19	2o-24	25-29	3o-34	35-39	4o-44	45-49	5o și peste	Total
Frecvențe	638	46.693	146.1o3	95.939	38.621	12.324	3.72o	3o2	29	344.369
%	o,2	13,6	42,2	27,8	11,2	3,6	1,1	o,1	x	1oo

Tabelul 8. Distribuția născuților vii în R.S.R., în 1982, după grupa de vîrstă a mamei. Sursa: Anuarul Statistic al RSR. 1983

2. Grafice atașate tabelelor de frecvențe

Forma fundamentală de prezentare a informației obținute în legătură cu o caracteristică este tabelul de frecvențe. Pentru o ilustrare mai sugestivă a mărimii frecvențelor, a raporturilor și diferențelor de populare a claselor, a modului cum se schimbă frecvențele cînd variabila trece de la valorile mici spre cele mari etc., tabelelor de frecvențe li se pot atașa diferite forme grafice x) Există foarte multe procedee de construcție a graficelor, chiar pentru o aceeași distribuție de frecvențe, dar acestea, în cea mai mare parte a lor, nu au decît o importanță relativ minoră, pur ilustrativă.

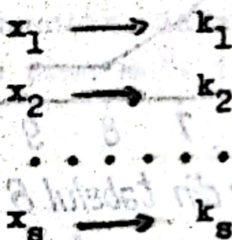
In cadrul rapoartelor de cercetare, însoțirea tabelelor de forme grafice este aproape obligatorie, deoarece beneficiarul poate astfel să înțeleagă mult mai ușor și mai rapid esența tabelelor. In studiile destinate publicațiilor științifice, în lucrările de diplomă sau în tezele de doctorat, folosirea figurilor gra-

x) Stricto sensu și tabelul este o formă de reprezentare grafică a

fice este recomandată doar în cazul unor tabele mai complexe, care presupun un efort mai mare spre a fi citite sau în situaţia unor variaţii interesante în rîndul frecvenţelor.

In cele ce urmează voi prezenta doar cîteva forme grafice principale, deoarece în legătură cu această problemă există o literatură destul de bogată. (A se vedea, de exemplu, Al.Bărbat, 1972). Forma fundamentală de prezentare grafică a datelor pentru o variabilă statistică este <u>poligonul frecvenţelor</u>.

Aşa cum am văzut la început, orice caracteristică este o funcţie de genul $f : \mathcal{A} \rightarrow \mathcal{K}$, ce aplică deci mulţimea claselor în mulţimea numerelor naturale. In cazul variabilelor statistice, mulţimea $\mathcal{A}$ devine o submulţime de <u>numere reale</u> şi deci o caracteristică de acest gen corespunde obişnuitelor funcţii studiate în şcoala generală şi liceu. Intr-adevăr, un tabel pentru o variabilă statistică X este o corespondenţă de forma:



$$\begin{array}{ccc} x_1 & \longrightarrow & k_1 \\ x_2 & \longrightarrow & k_2 \\ \cdots & \cdots & \cdots \\ x_s & \longrightarrow & k_s \end{array}$$

deci o mulţime de perechi ordonate (x_i, k_i), care se pot uşor reprezenta sub formă grafică.

Dacă luăm un sistem de axe de coordonate ortogonale, în care pe axa absciselor vom reprezenta valorile variabilei iar pe cea a ordonatelor frecvenţele, atunci mulţimea perechilor ordonate de mai sus primeşte drept imagine o mulţime de puncte în plan cu abscisele x_i şi ordonatele k_i. Pentru a se sugera trecerea de la o valoare x_i la valoarea următoare x_{i+1}, se unesc succesiv, două cîte două, aceste puncte cu segmente de dreptă, obţinîndu-se o linie poligonală ce reprezintă tocmai forma grafică menţionată: poligonul frecvenţelor.

De exemplu, distribuţiei din tabelul 6,i se va ataşa graficul din figura[1]. Dacă vom avea de a face cu intervale de grupare,

---

(continuare)
datelor. Cu toate acestea, noţiunile de grafic sau de reprezentare grafică s-au încetăţenit în limbajul curent de aşa manieră încît nu mai includ "tabelele" sau "prezentarea tabelară" a datelor.

pe axa absciselor se vor trece aceste intervale, iar verticalele vor fi ridicate din punctele de origine alese. Astfel, pentru distribuţia din tabelul 7 se poate construi un grafic ca cel din figura 2.

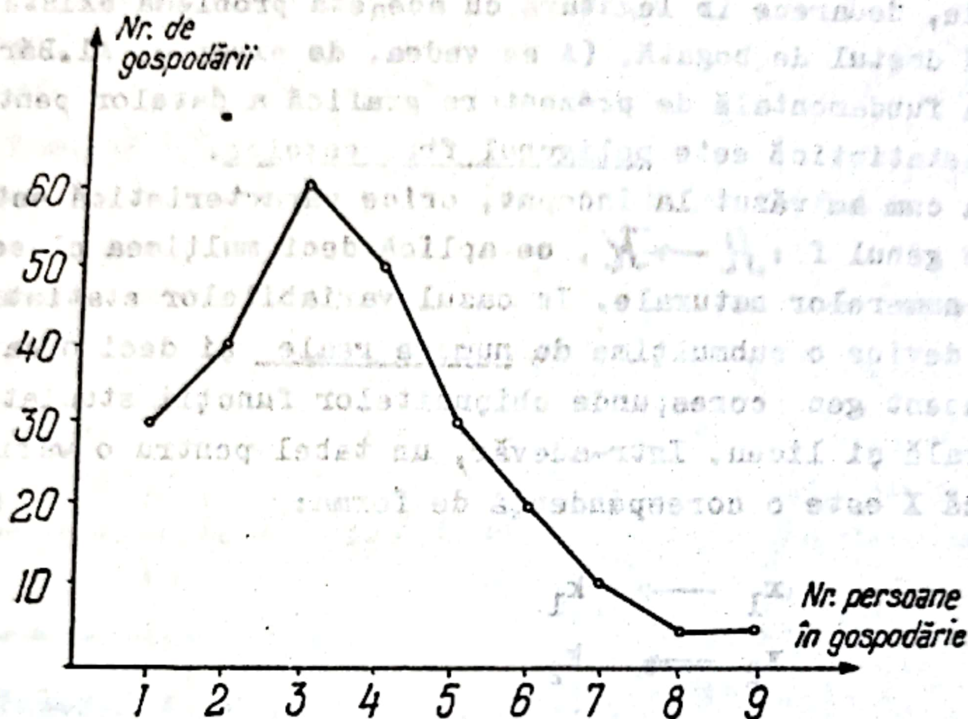


Fig. 1 Poligonul frecvenţelor din tabelul 6

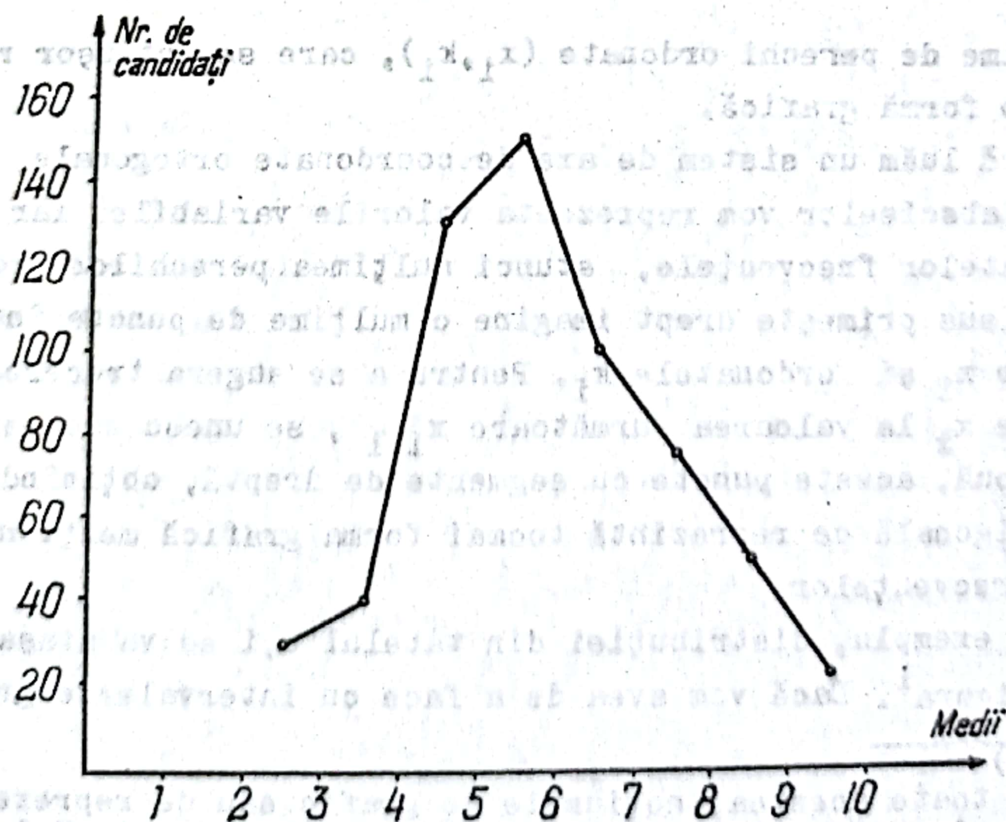


Fig. 2 Poligonul frecvenţelor din tabelul 7

Poligonul frecvenţelor - ca de altfel toate formele grafice ataşate distribuţiilor de frecvenţă - se poate construi fie utilizînd frecvenţele absolute, cum am procedat în fig.1 şi 2, fie cele relative, fie procentele. Acestea din urmă se recomandă mai ales acolo unde frecvenţele absolute prezintă variaţii foarte mari în rubricile tabelului. Dar utilizarea frecvenţelor relative şi procentelor devine indispensabilă dacă dorim să ilustrăm diferenţa de "comportament" dintre două populaţii diferite, în raport cu o aceeaşi variabilă. In astfel de cazuri, indiferent că vom trasa cele două linii poligonale pe aceleaşi sistem de axe sau vom folosi două grafice diferite, comparaţiile nu sînt posibile - aşa cum am mai spus - decît în termenii valorilor relative.

Să reprezentăm grafic şi datele tabelului 8, cu ajutorul procentelor (fig.3). Se observă că punctele de origine pentru intervalele semideschise de la capete s-au ales foarte aproape de limita superioară şi, respectiv, inferioară a lor, deoarece fenomenul urmărit nu se extinde prea mult în stînga şi în dreapta.

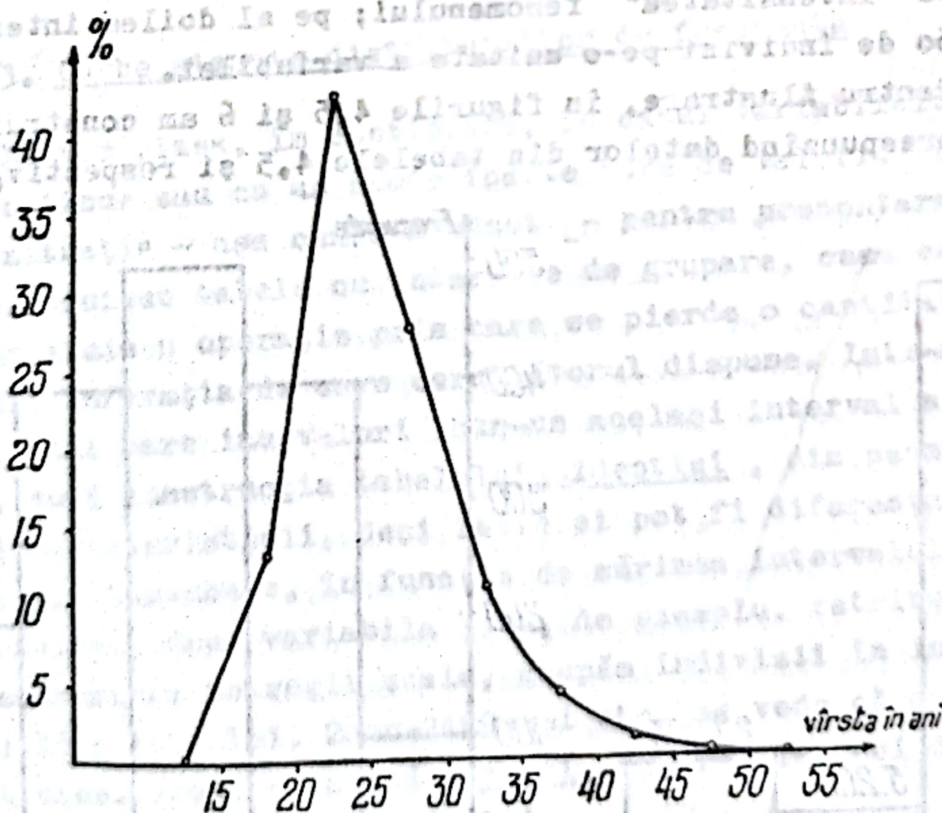


Fig. 3 Poligonul frecvenţelor din tabelul 8

A doua formă grafică fundamentală pentru distribuţiile de frecvenţă este histograma. Ea se aplică, de regulă, caracteristicilor calitative şi celor cantitative cu intervale de grupare cînd nu se alege un punct de origine pentru acestea. Principiul de construcţie este şi aici foarte simplu. Pe axa absciselor se reprezintă sub formă de segmente clasele caracteristicii sau intervalele de grupare, iar pe ordonată frecvenţele (absolute, relative, procente). Se ridică pe fiecare interval de pe abscisă un dreptunghi cu înălţimea indicată pe ordonată de reprezentarea frecvenţelor. Acest lucru este valabil, de fapt, numai cînd segmentele de pe abscisă sînt egale; în acest caz frecvenţele sînt reprezentate atît prin înălţimea dreptunghiurilor cît şi prin aria acestora. In cazul în care avem de a face cu intervale de grupare inegale, atunci bazele dreptunghiurilor variază în funcţie de lungimea acestor segmente iar frecvenţele pe intervale vor fi reprezentate numai de aria dreptunghiului. De pildă, dacă pe un interval de o unitate avem o frecvenţă de loo indivizi şi pe un altul de 2 unităţi aceeaşi frecvenţă, primul dreptunghi va avea înălţimea loo, pe cînd al doilea numai 5o. Aici înălţimea dreptunghiului reprezeintă "intensitatea" fenomenului; pe al doilea interval revin doar 5o de indivizi pe o unitate a variabilei.

Pentru ilustrare, în figurile 4,5 şi 6 am construit histograme corespunzînd datelor din tabelele 4,5 şi respectiv, 7.

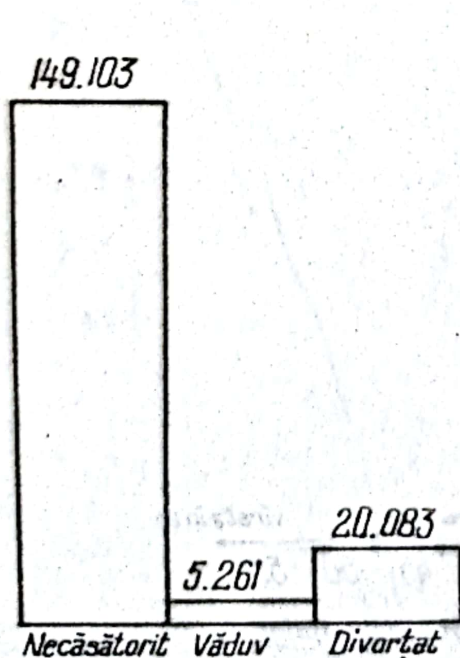


Fig. 4 Histograma frecvenţelor din tabelul 4

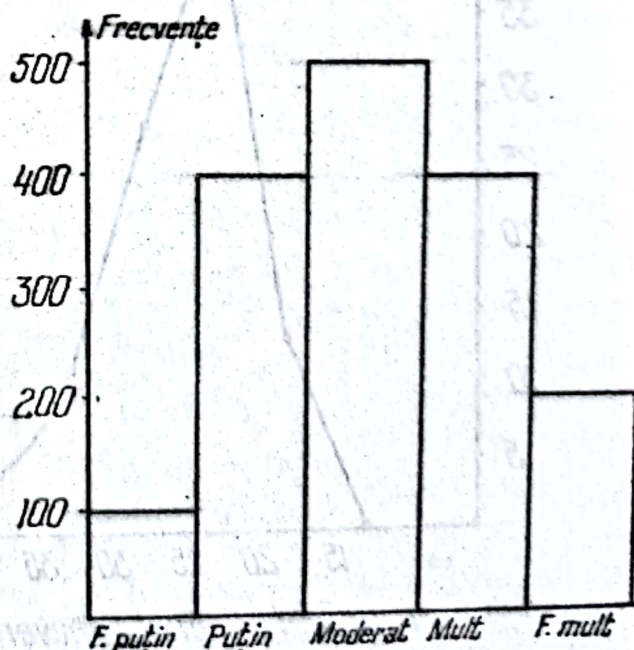


Fig. 5 Histograma frecvenţelor din tabelul 5

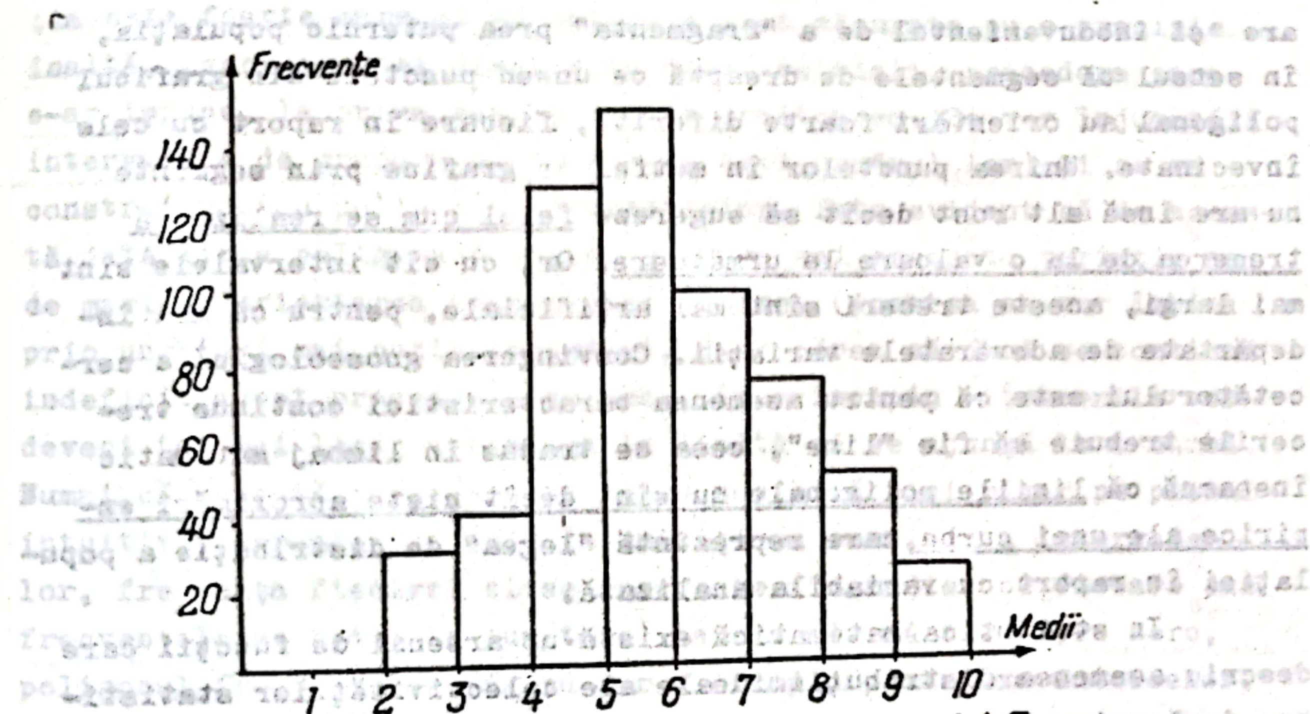


Fig.6 Histograma frecvențelor din tabelul 7

3. Curbe atașate distribuțiilor de frecvență

Ne vom plasa, în continuare, în cazul variabilelor statistice continue sau cu un număr foarte mare de valori. Intr-o asemenea situație - așa cum s-a văzut - pentru prezentarea datelor se construiesc tabele cu intervale de grupare, ceea ce, evident, constituie o operație prin care se pierde o cantitate însemnată din informația de care cercetătorul dispune. Intr-adevăr, toți indivizii care iau valori într-un același interval sînt considerați, după construcția tabelului, identici, din punctul de vedere al caracteristicii, deși între ei pot fi diferențe mai mult sau mai puțin însemnate, în funcție de mărimea intervalului de grupare. Astfel, dacă variabila este, de exemplu, retribuția și, pentru acoperirea întregii scale, grupăm indivizii în intervale de genul: 1500-1999 lei, 2000-2499 lei etc. se vede că doi indivizi sînt considerați ca avînd aceeași retribuție deși diferența între ei poate fi de aproape 500 lei.

Pe de altă parte, un număr redus de intervale de grupare - deși foarte favorabil pentru clasificări și analize ulterioare -

are şi inconvenientul de a "fragmenta" prea puternic populaţia, în sensul că segmentele de dreaptă ce unesc punctele din graficul poligonal au orientări foarte diferite, fiecare în raport cu cele învecinate. Unirea punctelor în astfel de grafice prin segmente nu are însă alt rost decît să sugereze felul cum se realizează trecerea de la o valoare la următoare. Or, cu cît intervalele sînt mai largi, aceste treceri sînt mai artificiale, pentru că mai îndepărtate de adevăratele variaţii. Convingerea gnoseologică a cercetătorului este că pentru asemenea caracteristici continue trecerile trebuie să fie "line", ceea ce tradus în limbaj matematic înseamnă că liniile poligonale nu sînt decît nişte aproximări empirice ale unei curbe, care reprezintă "legea" de distribuţie a populaţiei în raport cu variabila analizată.

In statistica matematică există un arsenal de funcţii care descriu asemenea distribuţii ideale ale colectivităţilor statistice, după mai multe tipuri de curbe. Dintre acestea, cea mai cunoscută - şi probabil cu aplicabilitatea cea mai largă - este aşa-numita curbă a lui Gauss, sub formă de clopot, cunoscută şi sub numele de curbă sau lege normală de distribuţie. Două probleme importante se pun în legătură cu trecerea de la un set de date individuale (valori pentru indivizi empiric) la asemenea curbe matematice:

Mai întîi, una practică, şi anume, avînd o distribuţie empirică, se urmăreşte să se vadă dacă se poate afirma că această distribuţie poate fi considerată o aproximaţie suficient de bună a uneia teoretice, mai exact trebuie testat dacă frecvenţele empirice pot fi privite cu particularizări ale celor teoretice. Se consideră că lotul de indivizi studiaţi reprezintă un eşantion reprezentativ al unei populaţii, potenţial infinite, iar testarea constă în calcularea semnificaţiei diferenţelor dintre frecvenţele empirice şi cele care ar rezulta, la acelaşi volum de populaţie şi la aceleaşi intervale de valori, dacă distribuţia urmează o curbă matematică ideală. Este deci o chestiune de testare a semnificaţiei diferenţelor, legată de eşantionare, despre care va fi vorba în ultima parte a cursului.

A doua problemă este una teoretică, de principiu, cerînd să se indice calea inversă, cea prin care s-ar putea trece de la valorile indivizilor empirici la o curbă. Presupunînd că popula-

are şi inconvenientul de a "fragmenta" prea puternic populaţia, în sensul că segmentele de dreaptă ce unesc punctele din graficul poligonal au orientări foarte diferite, fiecare în raport cu cele învecinate. Unirea punctelor în astfel de grafice prin segmente nu are însă alt rost decît să sugereze felul cum se realizează trecerea de la o valoare la următoare. Or, cu cît intervalele sînt mai largi, aceste treceri sînt mai artificiale, pentru că mai îndepărtate de adevăratele variaţii. Convingerea gnoseologică a cercetătorului este că pentru asemenea caracteristici continue trecerile trebuie să fie "line", ceea ce tradus în limbaj matematic înseamnă că liniile poligonale nu sînt decît nişte aproximări empirice ale unei curbe, care reprezintă "legea" de distribuţie a populaţiei în raport cu variabila analizată.

In statistica matematică există un arsenal de funcţii care descriu asemenea distribuţii ideale ale colectivităţilor statistice, după mai multe tipuri de curbe. Dintre acestea, cea mai cunoscută - şi probabil cu aplicabilitatea cea mai largă - este aşa-numita curbă a lui Gauss, sub formă de clopot, cunoscută şi sub numele de curbă sau lege normală de distribuţie. Două probleme importante se pun în legătură cu trecerea de la un set de date individuale (valori pentru indivizi empiric) la asemenea curbe matematice:

Mai întîi, una practică, şi anume, avînd o distribuţie empirică, se urmăreşte să se vadă dacă se poate afirma că această distribuţie poate fi considerată o aproximaţie suficient de bună a uneia teoretice, mai exact trebuie testat dacă frecvenţele empirice pot fi privite cu particularizări ale celor teoretice. Se consideră că lotul de indivizi studiaţi reprezintă un eşantion reprezentativ al unei populaţii, potenţial infinite, iar testare constă în calcularea semnificaţiei diferenţelor dintre frecvenţele empirice şi cele care ar rezulta, la acelaşi volum de populaţie şi la aceleaşi intervale de valori, dacă distribuţia urmează o curbă matematică ideală. Este deci o chestiune de testare a semnificaţiei diferenţelor, legată de eşantionare, despre care va fi vorba în ultima parte a cursului.

A doua problemă este una teoretică, de principiu, cerînd să se indice calea inversă, cea prin care s-ar putea trece de la valorile indivizilor empirici la o curbă. Presupunînd că popula-

ţia este foarte mare şi că valorile sînt măsurate cu o precizie înaltă, spre a putea discrimina bine indivizii, procedura care s-ar impune, la prima vedere, este următoarea. Se vor înjumătăţi intervalele de grupare (dublîndu-se deci numărul lor) şi se va construi un nou poligon al frecvenţelor. Este evident că de această dată linia poligonală obţinută nu va mai prezenta variaţii aşa de mari în orientarea segmentelor vecine. (Acestea se vor îmbina prin unghiuri mai puţin ascuţite). S-ar părea că dacă se continuă indefinit acest procedeu - evident, în principiu - trecerile vor deveni tot mai line, urmînd ca la limită să se ajungă la o curbă. Numai că această procedură nu este posibilă nici măcar pe planul intuitiv. Intr-adevăr, prin înjumătăţirea continuă a intervalelor, frecvenţa fiecărei clase descreşte (în medie, cu fiecare pas frecvenţele se reduc la jumătate) iar la limită tind spre zero, poligonul frecvenţelor căzînd pur şi simplu peste axa absciselor. Pentru a scăpa din această capcană a infinitului mare şi mic, trebuie mers pe o cale diferită, deschisă de utilizarea frecvenţelor sub o altă formă decît cea prezentată în tabelele anterioare: este vorba de aşa-numitele <u>frecvenţe cumulate.</u> Uneori această procedură este folositoare şi pentru scopuri empirice, aşa că ne vom opri puţin asupra ei.

S-a văzut că, în cazul unei variabile statistice, pentru fiecare valoare x_i s-a pus în corespondenţă o frecvenţă k_i, reprezentînd mulţimea indivizilor care primesc valoarea respectivă, după caracteristica studiată. Ne putem întreba uneori şi cîţi indivizi dintr-o populaţie iau valori <u>sub</u> sau <u>peste</u> o anumită valoare dată. Pentru a putea răspunde la întrebarea de mai sus, indiferent despre ce valoare ar fi vorba, tabelul repartiţiei de frecvenţă trebuie modificat, cumulînd toate frecvenţele <u>pînă la</u> fiecare x_i sau <u>peste</u> fiecare x_i. Se obţine astfel două tipuri de tabele de frecvenţe cumulate. De exemplu, pentru datele din tabelul 7, cumularea frecvenţelor va conduce la situaţia din tabelul 9.

Mediile (x_i)	Frecvente cumulate			
	A.Pînă la x_i		B.Peste x_i	
	Absolute	%	Absolute	%
2	0	0	6oo	1oo
3	3o	5,o	57o	95,o
4	7o	11,7	53o	88,3
5	2oo	33,3	4oo	66,7
6	35o	58,3	25o	41,7
7	45o	75,o	15o	25,o
8	525	87,5	75	12,5
9	575	95,8	25	4,7
1o	6oo	1oo	0	0

Tabelul 9. Frecvenţele cumulate ale caracteristicii prezentate în tabelul 7

Se observă imediat utilitatea tabelelor de acest gen. Ele oferă posibilitatea de a citi rapid cîţi indivizi sau ce procent dintre ei se află sub un anume prag sau îl depăşesc. De exemplu, în cazul de faţă 2oo din cei 6oo de candidaţi (33,3%) n-au reuşit să obţină media 5; sau 75 din 6oo (12,5%) au reuşit să depăşească media 8.

Observaţie. In tabelul 9 am trecut pe coloana întîi doar valorile, deşi caraczeristica era constituită prin intervale de grupare. Se observă, din felul cum au fost scrise aceste intervale, că de fapt porţiunea B a tabelului conţine frecvenţe cumulate reprezentînd numărul de indivizi care iau nu numai valori mai mari decît x_i dar şi pe x_i, inclusiv. Tot aşa de bine valorile întregi: 3,4,5... puteau constitui capetele superioare ale intervalelor şi atunci situaţia se inversa.

Prin reprezentarea grafică a frecvenţelor din tabelul 9, cu ajutorul poligonului frecvenţelor, se obţine figura 7.

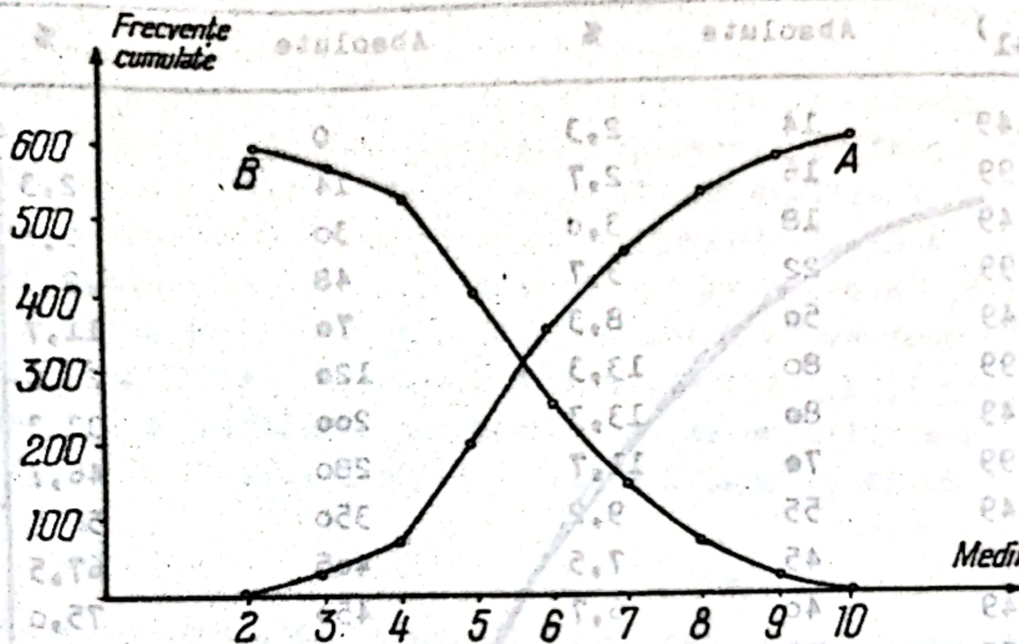


Fig. 7 Poligonul frecvențelor cumulate din tabelul 9

Este evident că întotdeauna frecvențele cumulate de tip A vor conduce la o linie poligonală ascendentă, în timp ce cele de tip B la una decendentă. Inclinația fiecărui segment față de axa orizontală este cu atît mai mare cu cît numărul indivizilor care se adaugă (sau se scad) de la o valoare la următoarea este mai mare. In cazul de față, unghiul cel mai mare cu axa Ox îl fac segmentele ce unesc punctele corespunzătoare valorilor 5 și 6, iar cel mai mic cele de la capetele liniei.

Mediile (x_i, x_{i+1})	Frecvenţe obişnuite		Frecvenţe cumulate, tip A	
	Absolute	%	Absolute	%
2 - 2,49	14	2,3	0	0
2,5o - 2,99	16	2,7	14	2,3
3,00 - 3,49	18	3,o	3o	5,o
3,5o - 3,99	22	3,7	48	8,o
4,00 - 4,49	5o	8,3	7o	11,7
4,5o - 4,99	8o	13,3	12o	2o,o
5,00 - 5,49	8o	13,3	2oo	33,3
5,5o - 5,99	7o	11,7	28o	46,7
6,00 - 6,49	55	9,2	35o	58,3
6,5o - 6,99	45	7,5	4o5	67,5
7,00 - 7,49	4o	6,7	45o	75,o
7,5o - 7,99	35	5,8	49o	81,7
8,00 - 8,49	3o	5,o	525	87,5
8,5o - 8,99	2o	3,3	555	92,5
9,00 - 9,49	15	2,5	575	95,8
9,5o - 1o	1o	1,7	59o	98,3
TOTAL	6oo	1oo	6oo	1oo

Tabelul 1o. Frecvenţele simple şi cumulate, obţinute prin înjumătăţirea intervalului de grupare (cazul prezentat în tabelul 7)

Revenind acum la problema curbelor de frecvenţe, se poate constata uşor că procedura de înjumătăţire a intervalelor nu ne mai conduce la impas dacă folosim frecvenţele cumulate. Să luăm în considerare numai tipul A şi să reluăm exemplul anterior, presupunînd că prin înjumătăţirea intervalelor se obţine datele din tabelul 1o, tabel în care s-au trecut şi frecvenţele cumulate de tip A (pînă la x_i, adică sub valoarea din stînga a intervalului).

In figura 8 am reluat, pentru comparaţie, linia poligonală A din figura 7, căreia i-am alăturat o linie poligonală construită cu frecvenţele cumulate din tabelul 1o. Se observă imediat cum unghiurile dintre segmentele vecine se "îndulcesc" şi cum, la li-

mită, linia poligonală tinde spre o curbă de genul celei trasate pe aceeaşi figură.

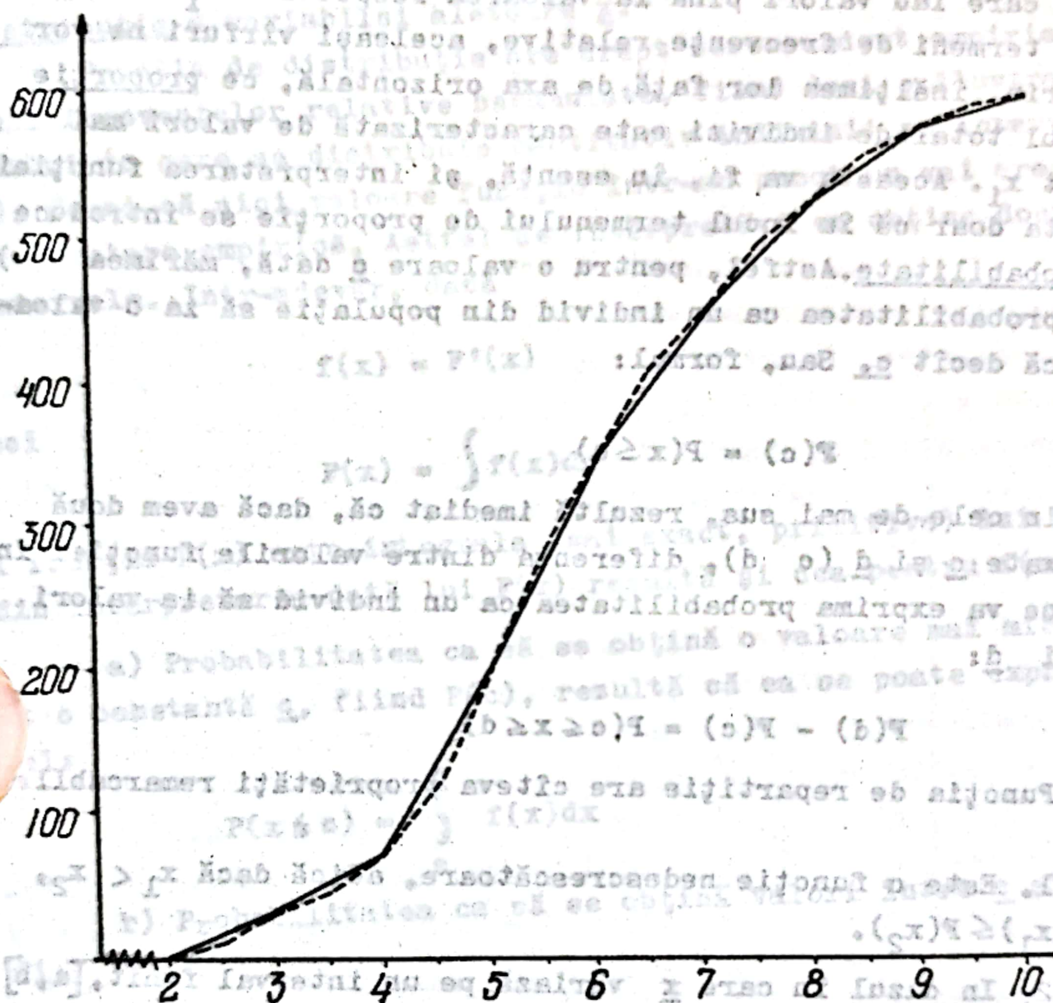


Fig. 8 Frecvenţele cumulate de tip A din Fig. 7 (linia continuă) şi cele obţinute prin înjumătăţirea intervalelor (linia punctată)

Operaţia descrisă mai sus se poate realiza atît cu frecvenţe absolute cît şi cu procente sau frecvenţe relative (la unitate). Să presuăunem că am utilizat aceste din urmă mărimi şi că am reuşit să facem saltul de la linii poligonale la o curbă. Am trecut atunci din domeniul statisticii descriptive în cel al matematicii, mai exact în cel al teoriei probabilităţilor. Curba obţinută va fi imaginea grafică a unei funcţii F(x), care poartă denu-

mirea de funcţie de repartiţie a unei variabile aleatoare X. Să vedem care este interpretarea acestei funcţii. Aşa cum am arătat, vîrfurile poligonului frecvenţelor cumulate ne indică volumul indivizilor care iau valori pînă la valoarea respectivă x_i. Dacă raţionăm în termeni de frecvenţe relative, aceleaşi vîrfuri ne vor indica, prin înălţimea lor faţă de axa orizontală, ce proporţie din numărul total de indivizi este caracterizată de valori mai mici decît x_i. Aceasta va fi, în esenţă, şi interpretarea funcţiei F(x), atîta doar că în locul termenului de proporţie se introduce cel de probabilitate. Astfel, pentru o valoare c dată, mărimea F(c) ne arată probabilitatea ca un individ din populaţie să ia o valoare mai mică decît c. Sau, formal:

$$F(c) = P(x \leq c)$$

Din cele de mai sus, rezultă imediat că, dacă avem două valori fixate c şi d (c d), diferenţa dintre valorile funcţiei în d şi c ne va exprima probabilitatea ca un individ să ia valori între c şi d:

$$F(d) - F(c) = P(c \leq x \leq d)$$

Funcţia de repartiţie are cîteva proprietăţi remarcabile evidente:

1. Este o funcţie nedescrescătoare, adică dacă $x_1 < x_2$, atunci $F(x_1) \leq F(x_2)$.

2. In cazul în care x variază pe un interval finit, [a,b],

$$F(a) = 0 \text{ şi } F(b) = 1$$

ceea ce corespunde faptului simplu că nici un individ nu ia valori mai mici decît a şi mai mari decît b.

3. Dacă intervalul de variaţie este $(-\infty, +\infty)$ atunci curba va tinde asimptotic la 0, spre $-\infty$, şi la 1, spre $+\infty$.

Vorbind despre linia poligonală a frecvenţelor cumulate, am observat că înclinarea acestora în raport cu axa Ox este în funcţie de proporţia indivizilor ce se adaugă de la o valoare la următoarea, respectiv de proporţia indivizilor cuprinşi în intervalul de grupare în cauză. Acest lucru se reflectă şi în cazul funcţiei F(x); dacă luăm două puncte de curbă şi le unim printr-o

coardă, la limită, coardele respective devin tangente la curbă, tangente care, aşa cum se ştie, indică "viteza de creştere" a funcţiei F(x) într-un punct. Se ajunge astfel imediat la o altă funcţie f(x), derivata lui F(x), care poartă numele de funcţie de distribuţie a variabilei aletoare X.

Funcţia de distribuţie are drept corespondent empiric poligonul frecvenţelor relative necumulate, fiind deci o ilustrare a modului în care se distribuie cantitativ indivizii pe intervalul (a,b). Numai că aici valoare funcţie într-un punct nu mai are nici o interpretare empirică. Astfel de interpretări se obţine doar pe subintervale. Intr-adevăr, dacă

$$f(x) = F'(x)$$

atunci

$$F(x) = \int f(x)dx$$

deci funcţia F(x) este integrala (mai exact, primitiva) lui f(x) şi din interpretarea dată lui F(x) rezultă şi cea pentru f(x):

a) Probabilitatea ca să se obţină o valoare mai mică decît o constantă c, fiind F(c), rezultă că ea se poate exprima şi astfel:

$$P(x \leq c) = \int_a^c f(x)dx$$

b) Probabilitatea ca să se obţină valori între c şi d va fi:

$$P(c \leq x \leq d) = \int_c^d f(x)dx$$

c) Cum F(b) = 1, rezultă că integrala lui f(x) pe tot intervalul (a,b) este egală cu unitatea:

$$\int_a^b f(x)dx = 1$$

relaţia fiind valabilă şi cînd limitele intervalului sînt infinite.

După cum se ştie, o interpretare grafică a integralei unei funcţii o constituie aria figurii delimitate de curbă, axa Ox şi două drepte verticale duse prin c şi d – limitele integralei. O asemenea situaţie este marcată în figura 9. Aria punctată reprezintă probabilitatea ca un individ să ia valori mai mici decît c, iar cea haşurată ca să ia valori între c şi d. Sau, vorbind în termeni empirici, aria punctată este egală cu proporţia indivizilor care nu depăşesc valoarea c, pe cînd cea haşurată proporţia sau ponderea celor cu valori între c şi d.

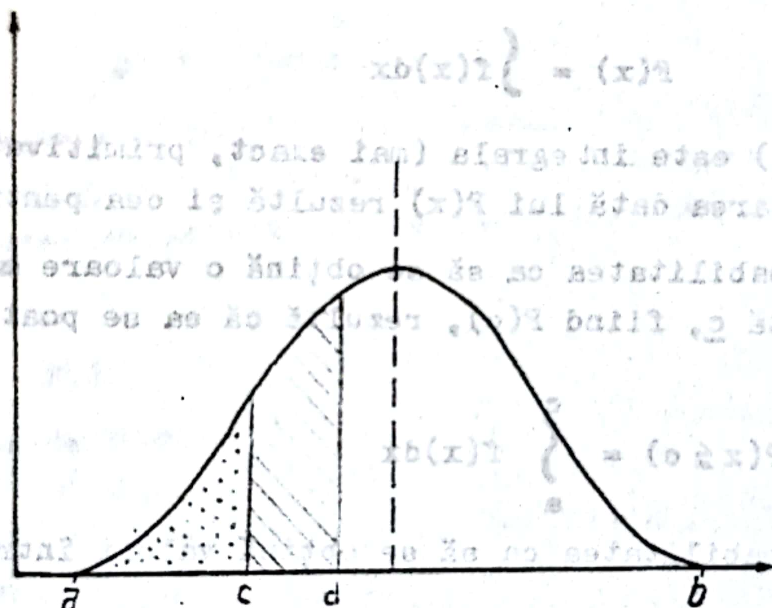


Fig. 9 O curbă simetrică

Diversele distribuţii empirice conduc spre cîteva modele teoretice de curbe, dintre care mai frecvent întîlnite sînt:

1. Curbele simetrice, ca cea din fig.9., în care maximum frecvenţelor se atinge la mijlocul intervalului (a,b) sau, dacă intervalul este (– , +), frecvenţele ating o zonă de maxim. scăzînd simetric spre stînga şi spre dreapta.

2. Curbele moderat asimetrice. Zona de frecvenţe maxime se află deplasată spre un capăt al intervalului şi apoi frecvenţele descresc cu viteze inegale într-o parte şi cealaltă.

3. Curbele în formă de J. Reflectă situaţiile în care maximum de cazuri se află într-un capăt al intervalului, descrescînd constant spre celălalt.

4. Curbele în formă de U. Descriu cazurile de distribuţii în care indivizii se concentrează spre cele două capete ale intervalului, avînd frecvenţe minime la mijlocul acestuia.

Dintre curbele simetrice, cea mai cunoscută este curba lui Gauss, definită pe intervalul ($-\infty$, $+\infty$). Experienţele empirice care pot duce la distribuţii normale sînt foarte multe. Amintesc doar două clasice. Să presupunem că cineva aruncă o bilă în intenţia de a o plasa pe o linie trasată la o oarecare distanţă. Măsurînd de fiecare dată distinţa de la punctul de impact al bilei cu solul la linia respectivă şi atribuind semnul " + " valorilor ce corespund unor aruncări lungi şi " - " celor scurte, va obţine o distribuţie de frecvenţe (indivizii fiind aruncările, iar valorile aceste distanţe cu semn) pe intervale de grupare, distribuţie care se va dovedi simetrică în jurul lui 0. O distribuţie asemănătoare se obţine şi în cazul unor măsurări repetate a lungimii unei bare cu un acelaşi instrument, suficient de fin pentru a putea evidenţia diferenţe foarte mici de măsurare. In ambele exemple amintite valorile variabilelor pot fi interpretate ca erori, de aruncare, respectiv, de măsurare. In general, se consideră că erorile care se datorează unor factori întîmplători, nesistematici, urmează o distribuţie normală în jurul lui zero, respectiv măsurătorile afectate doar de astfel de erori se distribuie normal în jurul valorii "adevărate" a obiectului de măsurat. De altfel, curba normală mai poartă denumirea şi de curbă a erorilor.

Desigur că nu toate curbele normale sînt identice; ele diferă în esenţă prin două elemente: valoarea în care se găseşte punctul central şi gradul de "împrăştiere" a indivizilor statistici în jurul acestei valori. De pildă, în primul caz amintit valoarea centrală va fi nulă iar, în al doilea, egală cu dimensiunea fizică de măsurat. De asemenea, în prima situaţie abaterile de la punctul de simetrie vor fi incomparabil mai mari decît în al doilea.

Datorită cîmpului foarte larg de aplicabilitate, virtuţile curbei normale sînt uneori prezentate exagerat, în sensul că

se consideră despre multe fenomene ca ar urma o asemenea distribuţie, fără a avea suficiente temeiuri pentru atare supoziţie. Astfel, în domeniul ştiinţelor sociale, se presupune adeseori, tacit sau explicit, că o serie de însuşiri ale populaţiilor umane sînt normal distribuite. Numai din faptul că valorile extreme sînt rare nu trebuie să tragem concluzia că distribuţia este simetrică şi, cu atît mai puţin, normală.

Voi încheia aceste sumare consideraţii asupra curbelor statistice cu precizarea că interesul faţă de ele nu constă numai sau în principal în funcţia lor ilustrativă. Dacă se constată că o distribuţie empirică poate fi suficient de bine apropiată de una teoretică atunci se pot extrapola asupra primeia toate cunoştinţele pe care le avem despre a doua. Dacă, în plus, prin cercetări repetate se constată mereu o asemenea apropiere, curba teoretică va reflecta ceea ce este constant în fenomene, o legitate sau regularitate socială, aflată în spatele unei mase de evenimente întîmplătoare.

Cap. VIII. INDICATORI AI CARACTERISTICILOR CANTITATIVE

Informaţia surprinsă sub forma unei variabile statistice, mai ales cînd numărul valorilor este mare, poate fi cu greu sesizată de cercetător sau de beneficiarul rezultatelor sale şi astfel se naşte nevoia concentrării ei sub forma unor mărimi unice, care să reflecte, sintetic, întreaga situaţie descrisă de o caracteristică. Desigur, o atare concentrare nu are numai rolul pragmatic sugerat aici, ci de cele mai multe ori, prin ea se realizează şi un scop gnoseologic major: mărimile unice calculate pentru caracteristica în cauză sînt chemate să reflecte ceea ce este general, tipic, normal, constant etc. în fenomenul social analizat. Mai mult, calculul acestor mărimi va constitui o condiţie necesară pentru continuarea analizelor la situaţii mai complexe, mai exact, pentru trecerea de la forma descriptiv-statistică la cea care ne ajută să explicăm fenomenele obiective descrise statistic.

In funcţie de genul de informaţie pe care o reprezintă, de ceea ce urmăresc să surprindă în cadrul unei distribuţii, mărimile respective, numite generic indicatori ai caracteristicilor, sînt de regulă, clasificate în mai multe categorii. Dintre acestea, în manualul de faţă, o să prezint doar două: categoria indicatorilor de poziţie şi cea a indicatori de împrăştiere.

1. Indicatori de poziţie

Clasa acestor indicatori cuprinde mărimile ce se calculează pentru caracteristicile cantiative, cu scopul de a surprinde, în mod foarte concentrat, nivelul valoric mijlociu al grupului de indivizi, Altfel spus, date fiind valorile luate de populaţia noastră, de la x_1 (cea mai mică) la x_s (cea mai mare), problema care se pune este de a determina o valoare care să caracterizeze cel mai bine ansamblul de indivizi, o valoare în jurul căreia se grupează celelalte, o valoare care să poată constitui un fel de etalon, de caz tipic pentru seria statistică a variabilei X. Atenţionez că o astfel de mărime poate coincide efectiv cu una dintre valorile

$x_1, \ldots, x_s$ luate de indivizii noştri sau nu va coincide. In orice caz, ea trebuie să se găsească undeva în interiorul intervalului (x_1, x_s) şi, de regulă, într-o poziţie centrală a acestuia.

Există evident mai multe posibilităţi de a construi un asemenea indicator, lucru vizibil şi din cerinţele pe care el este chemat să le îndeplinească, cerinţe formulate, recunosc, destul de vag în rîndurile precedente. Dar, cu tot caracterul vag al exprimării se poate vedea că nu toate cerinţele sînt echivalente. Astfel, de pildă, valoarea mijlocie poate să nu fie tipică, în sensul că ea nu caracterizează nici un individ din populaţie sau că în jurul ei nu găsim o concentrare deosebită a unităţilor noastre statistice. Astfel de situaţii se întîlnesc atunci cînd indivizii se plasează cu frecvenţe mai mari spre capetele intervalului de variaţie a valorilor (se distribuie după o curbă în formă de U), valoarea mijlocie fiind mai slab populată.

Dintre indicatorii de poziţie analizaţi în literatura statistică, voi prezenta doar trei, cei mai cunoscuţi şi mai des utilizaţi: media (aritmetică), mediana şi modul (valoarea modală) în ordinea în care i-am enumerat.

1.1. Media

După cum se ştie, există mai multe tipuri de mărimi care poartă denumirea de medie, dintre care cea mai răspîndită în analizele statistice este aşa-numita medie aritmetică. Dat fiind faptul că nu voi aborda şi celelalte genuri (media geometrică, media armonică), în cele ce urmează voi utiliza simplu denumirea de medie sau valoare medie, înţelegîndu-se că este vorba de cea aritmetică.

Noţiunea este foarte cunoscută pentru că este adesea folosită în viaţa de zi cu zi. Un student, care într-o sesiune ia, în ordinea examenelor date, următoarele note: 9, lo, 8, 8, lo , va spune imediat că a terminat sesiunea cu media 9, înlocuind prin această informaţie sintetică informaţia mai detaliată care s-ar fi transmis prin comunicarea celor 5 note; de asemenea, comunicîndu-se media 9, avem, implicit şi direct, o imagine generală asupra nivelului mijlociu la care a fost apreciată pregătirea sa.

Calculul mediei, în cazrui simple ca cel de mai sus, nu ridică, evident, nici o problemă; se adună notele (deci valorile caracteristicii) și se împarte la numărul cazurilor de intervin (numărul indivizilor), chit că în situația dată studentul nostru "vede" imediat media, folosind una din proprietățile acesteia.

In general, procedura formală de definire a indicatorului este următoarea. Avînd dată o mulțime de n indivizi, fiecare dintre aceștia luînd în raport cu o variabilă X o valoarea $x_i (i=1,2,...,n)$ atunci se definește media variabilei X, notată cu $\bar{x}$, prin relația:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

sau, prescurtat:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

Dacă populația statistică este mare , o aceeași valoare va fi luată de mai mulți indivizi și caracteristica noastră apare sub forma cunoscută a unei distribuții de frecvență. Atunci formula mediei se va scrie:

$$\bar{x} = \frac{k_1 x_1 + k_2 x_2 + ... + k_s x_s}{n}$$

sau, prescurtat:

$$\bar{x} = \frac{\sum_{i=1}^{s} k_i x_i}{n} \quad (2)$$

unde notațiile sînt cele din capitolul precedent: $x_1, x_2, ..., x_s$ sînt cele s valori distincte luate efectiv (sau, dacă variabila s-a construit prin intervale de grupare, cele s valori mijlocii ale fiecărui interval) iar $k_1, k_2, ..., k_s$ - frecvențele cu care sînt luate valorile respective.

Formula (1) poartă, în general, denumirea de medie aritmetică simplă, pe cînd (2), de medie ponderată. In fapt, este vorba de acelaşi lucru, deci nu sînt două medii de specii diferite, căci în (2) se concentrează în produse sume parţiale din (1). Intr-adevăr, dacă, de exemplu, lo indivizi statistici iau fiecare valoarea 15oo, atunci suma valorilor pentru ei se poate obţine adunînd pe 15oo cu el însuşi de lo ori sau mai simplu înmulţindu-l cu lo. Revenind la cazul studentului nostru, media notelor sale se poate calcula fie:

$$\bar{x} = \frac{9+1o+8+8+1o}{5} = 9$$

fie

$$\bar{x} = \frac{2.8+1.9 + 2.1o}{5} = 9$$

In domeniul sociologic, cele mai frecvente situaţii sînt cele care fac apel la formula (2), căci se lucrează cu populaţii mari. Pentru acest motiv, voi folosi mai frecvent în continuare formula (2), deşi, repet, între (1) şi (2) nu este decît o diferenţă de scriere: în primul caz, indicele i, după care se face însumarea, urmăreşte indivizii, adică valorile individuale scrise într-o anumită ordine şi în număr de n, pe cînd, în al doilea, valorile distincte după care se constituie cele s clase ale caracteristicii, aşezate, în general, de la cea mai mică la cea mai mare.

Pentru ilustrarea modului practic de calculare a mediei, să reluăm, mai întîi, caracteristica prezentată în capitolul precedent, în tabelul 6 (distribuţia gospodăriilor după numărul membrilor ce le compun)- In scopul facilităţii calculelor, datele scrise în tabele pe orizontală se trec acum pe verticală; se transpun cele două linii în coloane, una conţinînd valorile, alta frecvenţele. Se mai adaugă apoi o a treia coloană ce rezultă înmulţind fiecare frecvenţă cu valoarea corespunzătoare, ceea ce înseamnă că suma de la numărătorul formulei (2) se obţine adunînd cifrele de pe această din urmă coloană. Pentru cazul nostru, se va proceda astfel:

Valori x_i	Frecvenţe k_i	Produse $k_i x_i$
1	30	30
2	40	80
3	60	180
4	50	200
5	30	150
6	20	120
7	10	70
8	5	40
9	5	45
Total	250	915

Deci:

$$\bar{x} = \frac{915}{250} = 3,66$$

Adică mărimea medie a unei gospodării din localitatea respectivă este de 3,66 persoane. Ne aflăm aici într-un caz în care media nu coincide cu nici una din valori şi nici n-ar putea coincide decît dacă printr-o şansă deosebită, ar ieşi sub formă de număr natural. Chiar dacă este absurd să vorbim de o gospodărie cu 3,66 persoane, aceasta nu ne împiedică deloc să folosim mărimea medie pentru a indica nivelul mijlociu al domensiunii gospodăriilor din satul studiat.

Observaţie. Suma $\sum k_i x_i$, adică suma valorilor tuturor unităţilor din populaţia studiată, poate uneori avea interpretări foarte directe şi ea poate fi cunoscută fără a recurge la calculele de mai sus. De pildă, în exemplul de faţă, 915 nu este decît numărul total de indivizi ai satului, ce putea fi obţinut fără a construi propriu-zis o caracteristică; deci şi media poate fi uneori aflată fără detalierea datelor. Dacă, un alt caz, ştim că într-o întreprindere fondul de retribuţie pe o lună este de 1.500.000 lei şi întreprinderea are 500 de angajaţi, este evident că retribuţia medie pe lucrător va fi de 3.000 lei, fără a şti exact

Valori x_i	Frecvențe k_i	Produse $k_i x_i$
1	3o	3o
2	4o	8o
3	6o	18o
4	5o	2oo
5	3o	15o
6	2o	12o
7	1o	7o
8	5	4o
9	5	45
Total	25o	915

Deci:

$$\bar{x} = \frac{915}{25o} = 3,66$$

Adică mărimea medie a unei gospodării din localitatea respectivă este de 3,66 persoane. Ne aflăm aici într-un caz în care media nu coincide cu nici una din valori și nici n-ar putea coincide decît dacă printr-o șansă deosebită, ar ieși sub formă de număr natural. Chiar dacă este absurd să vorbim de o gospodărie cu 3,66 persoane, aceasta nu ne împiedică deloc să folosim mărimea medie pentru a indica nivelul mijlociu al domensiunii gospodăriilor din satul studiat.

Observație. Suma $\sum k_i x_i$, adică suma valorilor tuturor unităților din populația studiată, poate uneori avea interpretări foarte directe și ea poate fi cunoscută fără a recurge la calculele de mai sus. De pildă, în exemplul de față, 915 nu este decît numărul total de indivizi ai satului, ce putea fi obținut fără a construi propriu-zis o caracteristică; deci și media poate fi uneori aflată fără detalierea datelor. Dacă, un alt caz, știm că într-o întreprindere fondul de retribuție pe o lună este de 1.5oo.ooo lei și întreprinderea are 5oo de angajați, este evident că retribuția medie pe lucrător va fi de 3.ooo lei, fără a ști exact

cum se distribuie retribuţiile efective.

In situaţiile în care produsele $k_i x_i$ duc la valori mari sau presupun calcule mai greoaie din pricina zecimalelor, se poate aplica o formulă de calcul simplificat a mediei. Se pleacă de la constatarea că dacă din fiecare valoare a caracteristicii se scade o mărime constantă şi se calculează media cu diferenţele rămase, pentru a se reveni la media adevărată trebuie să i se adauge primeia valoarea constantă scăzută. Sau, scris riguros:

$$\bar{x} = \frac{\sum_{i=1}^{s} k_i(x_i - a)}{n} + a \qquad (3)$$

unde a este constanta menţionată. Se poate arăta uş r că formulele (2) şi (3) sînt echivalente. Să aplicăm formula (3) pentru situaţia prezentată în tabelul 7 din capitolul anterior. Era vorba acolo de distribuţia numărului de candidaţi la admitere pe intervale de note: 2 - 2,99 , 3 - 3,99 , ..., 9 - 1o. Să ataşăm fiecărui interval drept valoare caracteristică notele 2,5o, 3,5o,... ...,95o, care devin astfel valorile variabilei statistice. Pentru a evita înmulţirea cu zecimale, vom alege ca valoare ajutătoare:

$$a = 5{,}5o$$

Atunci, calculul se ordonează astfel:

x_i	k_i	$x_i - a$	$k_i(x_i - a)$ +	$k_i(x_i - a)$ −
2,5o	3o	-3		9o
3,5o	4o	-2		8o
4,5o	13o	-1		13o
5,5o	15o	0	0	
6,5o	1oo	1	1oo	
7,5o	75	2	15o	
8,5o	5o	3	15o	
9,5o	25	4	1oo	
Total	6oo		5oo - 3oo = 2oo	

După cum se vede, în ultima rubrică am aranjat cifrele pe două coloane: cele pozitive în stînga și cele negative în dreapta, pentru a facilita însumarea algebrică. S-a obținut că suma de la numărătorul fracției este 2oo, valoare ce împărțită la cei 6oo de indivizi ne va da o,33. Deci

$$\bar{x} = 0,33 + 5,50 = 5,83$$

Aceasta este deci nota medie obținută de candidații la admitere, în cazul ipotetic prezentat.

Observație. In cazul caracteristicilor cu intervale de grupare, media care se obține este doar o aproximație a mediei adevărate, pentru calculul căreia ar fi trebuit luată în considerate valoarea exactă a fiecărui individ. De regulă, aproximația astfel obținută este satisfăcătoare, deoarece erorile ce se introduc prin alegerea valorilor caracteristice pentru intervale se compensează de-a lungul scalei intervalelor. Astfel, dacă distribuția este simetrică sau moderat asimetrică, pe intervalele din stînga indivizii au tendința de a se grupa mai mult spre capătul din dreapta (deci mijlocul intervalului subestimează media valorilor din interval), pe cînd în partea superioară a scalei, datorită descreșterii frecvențelor, valorile caracteristice alese supraestimează mediile parțiale pe intervale.

Proprietăți ale mediei

1. In calitatea sa de mărime mijlocie, este evident că media trebuie să se găsească plasată pe scala valorilor între cea mai mică și cea mai mare valoare:

$$x_1 < \bar{x} < x_s$$

Proprietatea poate primi ușor o demonstrație riguros matematică.

2. Media este o mărime exprimată în aceleași unități de măsură ca și variabila pe care o reprezintă. Astfel, dacă valorile variabilei sînt note, ani, unități monetare etc., media va ilustra același gen de mărimi.

3. Calitatea de mărime mijlocie a mediei poate fi exprimată și prin proprietatea că suma abaterilor valorilor individuale

de la medie este nulă. Sau, altfel spus, abaterile valorilor din stînga mediei se compensează cu ale celor din dreapta ei. Formal proprietatea poate fi astfel exprimată:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

pentru cazul în care se aplică formula (1) sau:

$$\sum_{i=1}^{s} k_i(x_i - \bar{x}) = 0$$

pentru formula (2).

Intr-adevăr, dacă luăm, de pildă, relaţia a doua şi desfacem parantezele, egalitatea devine exchivalentă cu:

$$\sum_{i=1}^{s} k_i x_i - \sum_{i=1}^{s} \bar{x}.k_i = 0$$

Sau, dînd factor comun pe $\bar{x}$ în termenul al doilea:

$$\sum_{i=1}^{s} k_i x_i - \bar{x} \sum_{i=1}^{s} k_i = 0$$

Impărţind fiecare termen al relaţiei cu $\underline{n}$, rezultă:

$$\frac{\sum_{i=1}^{s} k_i x_i}{n} - \frac{n.\bar{x}}{n} = 0$$

Primul termen al diferenţei este exact valoarea medie dată de (2), iar în termenul care se scade, după simplificarea cu $\underline{n}$, rămîne acelaşi lucru:

$$\bar{x} - \bar{x} = 0$$

şi formula este demonstrată.

4. Tot ca o expresie a calităţii de mărime mijlocie, media posedă şi proprietatea că, înmulţită cu numărul indivizilor, se obţine exact suma valorilor individuale. Faptul rezultă imediat dacă formulele (1) şi (2) se înmulţesc cu n:

$$\sum_{i=1}^{n} x_i = n.\bar{x}$$

$$\sum_{i=1}^{s} k_i x_i = n.\bar{x}$$

Proprietatea reflectă faptul că media este acea valoare pe care ar lua-o indivizii dacă toţi ar fi identici din punctul de vedere al caracteristicii respective. De pildă, suma notelor studentului despre care am pomenit la început este de 45, sumă la care s-ar fi ajuns dacă studentul în cauză ar fi luat la toate cele cinci examene nota 9. Sau, în alt exemplu folosit, dacă toţi cei 5oo de lucrători ar fi avut aceeaşi retribuţie - 3.ooo lei - cît reprezintă media pe întreprindere, fondul total de retribuţie ar fi acelaşi.

5. Media aritmetică rămîne neschimbată dacă frecvenţele se înmulţesc sau se împart cu un acelaşi număr. Proprietatea este evidentă, căci amplificînd fiecare frecvenţă k_i cu o valoarea m, aceasta din urmă poate fi dată factor comun la numărător; la fel şi numărul indivizilor de la numitor (care nu este decît suma frecvenţelor) se va înmulţi cu m şi, prin urmare, fracţia se va simplifica imediat cu m. Această proprietate ar putea fi utilă în calcule în situaţiile cînd frecvenţele au un multiplu comun ce poate fi dat factor şi deci cu care se va simplifica fracţia. Din păcate, în calculele pe date reale, astfel de situaţii sînt improbabil de întîlnit.

In contextul acestei proprietăţi, mai menţionez că media aritmetică poate fi scrisă şi cu ajutorul frecvenţelor relative. Impărţind fiecare termen de la numărătorul fracţiei din formula (2) cu n şi deci eliminînd numitorul, se ajunge la:

$$\bar{x} = \sum_{i=1}^{s} p_i x_i \qquad (2')$$

6. Dacă înmulţim sau împărţim valorile caracteristicii cu acelaşi număr, valoarea medie se înmulţeşte sau se împarte şi ea tot cu numărul respectiv. Proprietatea rezultă şi ea imediat tot din formula de definiţie a mediei. Coroborînd această proprietate cu cea privind translaţia valorilor, folosită în formula (3) de simplificare a calculelor, ajungem la concluzia că dacă valorile variabilei X suferă o transformare lineară, atunci media noii variabile, X', se regăseşte prin aceeaşi transformare. Adică, dacă

$$X' = a + b.X$$

atunci

$$\bar{x}' = a + b.\bar{x}$$

După cum se ştie, media este o mărime foarte curent folosită în diferite domenii ale cunoaşterii. Uneori utilizarea sa este abuzivă, mai cu seamă atunci cînd ea este luată drept valoare mijlocie sau chiar tipică şi, de fapt, nu se întîmplă acest lucru. Intr-adevăr, din formula (2') se vede foarte bine că media este în funcţie de doi factori deosebiţi: valorile caracteristicii şi ponderile pe care le au indivizii ce iau aceeaşi valoare în populaţie. Se poate întîmpla ca asupra mărimii medii să aibă o influenţă sensibilă una sau cîteva valori foarte mari ori foarte mici, dar care caracterizează un număr restrîns de unităţi ale populaţiei. Problema se va înţelege mai bine dacă dăm un exemplu foarte simplu. Să presupunem că într-o echipă de muncă lucrează patru muncitori necalificaţi care, în funcţie de vechimea lor primesc o retribuţie lunară de 25oo, 26oo, 27oo şi respectiv 28oo lei. Alături de ei se mai află şi un maistru cu o retribuţie lunară de 44oo lei. Făcînd media retribuţiei pe întreaga echipă se va obţine:

$$\bar{x} = \frac{2500+2600+2700+2800+4400}{5} = \frac{15.000}{5} = 3000$$

Deci, dacă, în medie, un membru al echipei respective cîştigă 3ooo lei lunar, în realitate patru din cinci se află sub această valoare şi doar unul singur (mult) peste ea. In asemenea cazuri, media nu redă evident o valoare mijlocie a grupului, în sensul intuitiv al termenului de mijlociu, reflectat mai bine de un alt indicator: mediana.

Comentînd situaţii asemănătoare cu cea de mai sus, autorii unui manual de statistică dedicat economiştilor (Vezi M.Biji şi E.Biji, 1979, pg.lol) merg pînă acolo încît recomandă calcularea mediei doar pentru situaţii omogene, tocmai pentru a se evita interpretări greşite. Cred că se exagerează aici puţin; din cauza unor utilizări şi interpretări greşite nu se poate ajunge la a interzicee calcularea unui indicator.

Important pentru cel care foloseşte indicatorii statistici, media în cazul de faţă, este să înţeleagă semnificaţia exactă a valorii calculate şi să o utilizeze exclusiv în scopurile care se potrivesc acestei semnificaţii. Mediile nu au un caracter fictiv, nu sînt lipsite de conţinut, nu sînt neprezentative, în situaţii ca cea de mai sus, cum lasă să se înţeleagă autorii menţionaţi. Ele pur şi simplu nu ne dau ceea ce unii aşteaptă de la ele sau pretind că ne dau. Cerinţa omogenităţii are desigur o anumită raţiune. Media, am spus, nu poate să reflecte decît o parte din informaţia surprinsă în caracteristică şi este evident că, cu cît populaţia este mai omogenă, cu atît media va "trăda" mai puţin din cantitatea de informaţie. La limită (vezi proprietatea 4 de mai sus), dacă toţi indivizii iau aceeaşi valoare, aceasta va fi şi media, deci indicatorul nostru va reda <u>întreaga</u> informaţie conţinută în caracteristică. Faptul că anumite situaţii reale sînt departe de acest caz limită nu trebuie să ne ducă la interzicerea calculării mediei, ci la necesitatea de a determina, pe lîngă acest indicator de poziţie, şi alţii, în speţă cei de împrăştiere şi de asimetrie a distribuţiilor. Dacă însă dorim să reflectăm mai clar valorile mijlocii sau tipice, vom alege dintre indicatorii de poziţie pe cei potriviţi pentru asta. Unul dintre ei este tocmai cel despre care vom vorbi în continuare.

1.2. Mediana

Mediana, notată Me, reprezintă valoarea ce corespunde individului care împarte în două părţi egale colectivitatea statistică ai cărei membri sînt presupuşi a fi ordonaţi, în funcţie de valorile luate.

O primă precizare se impune imediat. Mediana este o valoare şi ea nu trebuie confundată cu individul median sau, cu altă denumire, cu unitatea mediană a populaţiei, reprezentînd individul din populaţie care ia valoarea mediană. Să exemplificăm definiţia cu un exemplu foarte simplu şi la îndemînă. Să presupunem că un student la filosofie are într-o sesiune cinci examene, pe care le susţine în ordinea următoare şi le trece cu notele menţionate:

Logică	- 6
Filosofie	- 9
Istorie universală	- 8
Istoria României	- 1o
Economie politică	- 6

Ordonînd mulţimea celor 5 unităţi care formează colectivitatea noastră, de la cea cu valoare minimă la cea cu valoare maximă, obţinem:

Unităţi	Rang
Logică	1-2
Ec.politică	1-2
Ist.universală	3
Filosofie	4
Ist.României	5

Conform definiţiei şi precizări de mai sus, unitatea mediană va fi "Istoria universală" care ocupă rangul mijlociu în ordonarea efectuată, în sensul că înaintea ei sînt două unităţi şi după ea alte două. Valoarea mediană, fiind valoarea individului median, ea este, în cazul nostru, nota 8.

Me = 8

Prin urmare, pentru determinarea medianei trebuie efectuate următoarele operaţii: ordonăm indivizii din populaţie după valorile luate (ceea ce se realizează, de regulă, atunci cînd se construieşte variabila statistică), determinăm rangul individului median şi apoi valoarea ce corespunde acestuia. Am văzut că pentru 5 indivizi, rangul unităţii mediane este 3. In general, pentru n indivizi,

$$\text{rangul individului median} = \frac{n+1}{2}$$

De pildă, într-o populaţie cu 999 elemente componente, individul median va avea rangul 5oo (1ooo : 2)

Se observă că prin formula de mai sus se obţine o valoare întreagă numai dacă seria statistică este definită pe un număr impar de indivizi. Intr-adevăr, numai atunci există realmente un individ median. In cazul cînd n este un număr cu soţ, individul median este unul imaginar, al cărui loc se găseşte între indivizii de rang n/2 şi n/2+1. De pildă, dacă studentul ale cărui rezultate le-am ilustrat mai sus ar mai fi avut un examen, luat cu nota lo, seria valorilor ar fi:

6 , 6 , 8 , 9 , lo , lo

locul individului median imaginar

Dacă există un individ median real, atunci nu există nici o valoare mediană propriu-zisă. Totuşi pentru a nu reduce generalitatea aplicabilităţii indicatorului, se defineşte valoarea mediană şi pentru n cu soţ. Ea este media valorilor pe care le iau indivizii de rang n/2 şi n/2 + 1. In situaţia de mai sus:

$$Me = \frac{8+9}{2} = 8{,}5o$$

Să vedem , în continuare, cum se rezolvă problema determinării medianei în cazul general, cînd numărul indivizilor este mare şi avem de a face cu caracteristici de genul celor prezentate în capitolul anterior. Vom distinge două situaţii: cînd valorile

caracteristicii sînt discrete şi efectiv luate de indivizii din fiecare clasă, pe de o parte, şi cînd se lucrează cu intervale de grupare, pe de alta.

Pentru primul gen de caracteristici, nu se pune nici o problemă deosebită, rezolvarea fiind identică cu cea din cazurile simple prezentate deja. Dacă numărul indivizilor este fără soţ, se va putea identifica rangul unităţii mediane, care se află într-o clasă caracterizată de o valoare unică, valoarea ce va reprezenta mediana. Dacă numărul unităţilor este cu soţ, se determină rangurile n/2 şi n/2 + 1. Există atunci două posibilităţi: cei doi indivizi, corespunzînd acestor ranguri, se află în aceeaşi clasă şi atunci mediana este valoarea clasei respective., sau primul se află într-o clasă, caracterizată de valoarea x_i şi al doilea în clasa următoare, x_{i+1}. In această din urmă situaţie (foarte improbabilă, de altfel) se va efectua media valorilor celor două clase, ţinîndu-se cont de efectivele fiecăreia. Adică:

$$Me = \frac{k_i x_i + k_{i+1} x_{i+1}}{k_i + k_{i+1}}$$

Reluînd exemplul din capitolul anterior cu dimensiunea gospodăriilor (tabelul 6), observăm că mediana trebuie să se găsească între gospodăria cu rangul 125 şi 126, deoarece totalul de gospodării este un număr cu soţ: 250. Pentru a ajunge la rangurile respective, însumăm succesiv frecvenţele:

30 + 40 = 70

70 + 60 = 130

Ne oprim aici căci am depăşit cifra 126, şi, prin urmare, ambele gospodării căutat se află în clasa cu valoarea 3. Deci:

$$Me = 3$$

sau, altfel spus, gospodăria mijlocie a satului este formată din 3 persoane.

Ca să nu mai construim un exemplu nou, să presupunem, pentru a ilustra cealaltă situaţie, că mai adăugăm 10 gospodării celor 250, toate posedînd un număr mare de persoane (de pildă ,

5 cu 8 membri şi alte 5 cu 9). Atunci totalul va ajunge la 26o şi mediana se va găsi între cea cu rang 13o şi 131. Dar, aşa cum am văzut, gospodăria cu rang 13o încheie şirul celor cu dimensiune 3 şi deci următoarea va cădea în clasa celor cu dimensiune 4. Cum cele două clase nu au aceleaşi frecvenţe, nu vom putea spune că mediana este 3,5o , ci vom lua în considerare numărul de apariţii pentru cele două valori.
Sau, cu altă exprimare, se efectuează media ponderată a valorilor 3 şi 4:

$$Me = \frac{6o.3 + 5o.4}{6o + 5o} = \frac{38o}{11o} = 3,45$$

In fine, rămîne să elucidăm situaţia caracteristicilor realizate prin intervale de grupare. Aici ne vom opri doar la cazul cînd individul median cade într-un interval, deoarece dacă este la limita dintre două intervale - în cazul cînd <u>n</u> este cu soţ, căci altfel individul median e determinat şi se va afla sigur într-un interval - , putem fie să atribuim medianei valoarea care desparte cele două intervale, fie să procedăm ca mai sus, utilizînd valorile caracteristice intervalelor. Deci în situaţia cea mai frecvent întîlnită, individul median se va găsi printre cei care constituie frecvenţa unui interval de valori, care se va numi <u>interval median.</u> De regulă, interesul cercetătorului este satisfăcut odată cu determinarea intervalului median. De exemplu, privind caracteristica redată în tabelul 7 din capitolul precedent (notele la admitere), se observă imediat că, dat fiind totalul de 6oo , individul median se află în intervalul de note de la 5 la 5,99, care deci redă nivelul mijlociu al notelor obţinute.

Dacă din diferite raţiuni dorim să mergem mai departe şi să calculăm <u>o valoare mediană</u>, în situaţiile de mai sus, atunci se poate proceda în felul următor. Cînd avem informaţii detaliate asupra valorilor fiecărui individ, putem urmări situaţia indivizilor din intervalul median şi să-l determinăm pe cel care ocupă poziţia de mijloc. De pildă, în cazul mai sus invocat, individul median va fi între cel cu rang 3oo şi 3o1. Dar cum pînă la intervalul median sînt 2oo indivizi, rămîne doar să găsim, dintre cei 15o cuprinşi în intervalul median, pe cel cu rang între 1oo (3oo-2oo) şi 1o1.

In lipsa informaţiilor detaliate despre valorile fiecărui individ din intervalul median, putem proceda - pe baza supoziţiei că indivizii se distribuie uniform pe acest interval - la un calcul simplu de interpolare. In cazul cu notele, am calculat că pînă la valoarea 5 sînt 2oo indivizi şi deci pentru a ajunge la mediană trebuie să mai adăugăm loo (considerăm că individul median este cel cu rang 3oo, căci oricum calculul este aproximativ). Atunci aplicăm regula de trei simplă:

Dacă la 15o indivizi din intervalul median corespunde o lungime a intervalului de 1 notă, atunci la loo indivizi va corespunde x. Si:

$$x = \frac{100.1}{150} = 0,67$$

Adăugînd o,67 la 5, se va obţine:

$$Me = 5,67$$

Utilizarea medianei este frecventă şi binevenită dacă se urmăreşte exclusiv indicarea unor mărimi mijlocii, în situaţii în care distribuţiile sînt puternic asimetrice şi deci media este afectată de valori mari sau mici. Din păcate, lipsa unei formule de calcul exprimabilă algebric într-un mod foarte simplu şi unitar - ca în cazul mediei - face ca mediana să se găsească departe în urma mediei în întrebuinţările statistice. Ea este aproape total evitată atunci cînd trebuie să intre ca element parţial în calculul altor indicatori statistici derivaţi.

1.3. Modul

Modul (şi nu modulul, cum greşit mai apare uneoir chiar şi în manualele de statistică) sau valoarea modală sau dominanta este acea valoare din seria caracteristicii cantitative care este luată de cel mai mare număr de indivizi, deci valoarea căreia îi corespunde frecvenţa maximă. Valoarea modală se foloseşte pentru a ilustra cazurile numite tipice în populaţia studiată, cazurile cele mai frecvente. Este limpede că determinarea acestui indicator

se va face numai pe o masă mare de indivizi, din care se poate degaja o eventuală concentrare a indivizilor pe sau în jurul unei valori date. De pildă, în exemplul cu gospodăriile, se observă că cea mai mare frecvenţă apare sub valoarea 3, ceea ce înseamnă că în localitatea studiată gospodăria modală (sau tipică) este cea cu 3 persoane.

Determinarea valorii modale se face, după cum se vede, simplu, prin inspecţia vizuală a tabelului caracteristicii. Singura problemă apare în cazul intervalelor de grupare cînd, evident, va trebui în primul moment determinat intervalul modal, adică cel în care frecvenţa este maximă. Aşa de exemplu, în situaţia candidaţilor la admitere, intervalul modal al notelor este cel de la 5 la 6. Se poate şi aici imagina - ca şi pentru mediană - o procedură de calcul a valorii modale. Personal înclin să cred că o astfel de valoare calculată pentru intervale de grupare este prea puţin semnificativă pentru caracteristicile ce redau o situaţie empirică, deoarece aici efectiv populate sînt intervalele şi nu valorile acestora. Desigur, dacă anumite calcule reclamă utilizarea unei valori modale, atunci trebuie utilizate formule pentru determinarea acesteia în cadrul intervalului modal.

2. Indicatori de împrăştiere

Indicatorii din această clasă se ataşează, de regulă. indicatorilor de poziţie pentru a îmbogăţi informaţia asupra caracteristicii studiate. Ei urmăresc, în esenţă, să măsoare gradul de împrăştiere sau dispersia indivizilor în cadrul scalei de valori pe care le iau. Tradusă problema în alţi termeni, se vede că este vorba de a determina gradul de omogenitate sau de eterogenitate al unei populaţii în raport cu o variabilă statistică. Sau - o altă interpretare posibilă, deşi mai puţin uzuală, dar deloc lipsită de importanţă - este cea în termeni de inegalitate între indivizi, în raport cu factorul social tradus în caracteristică. O dispersie mare este reflectarea unui grad înalt de inegalitate, pe cînd una mică, specifică populaţiilor omogene, ne arată, evident, un grad redus de inegalitate între indivizii populaţiei.

La prima vedere, se pare că problema dispersiei poate fi

ușor expediată urmărind întinderea scalei de valori a caracteristicii. De pildă, dacă în două clase de elevi dăm cîte un extemporal la limba română și, în prima clasă, notele obținute vor fi, să zicem, numai între 7 și 9, iar în a doua între 4 și 1o, vom putea, bineînțeles, să ne exprimăm că aceasta din urmă este mai eterogenă, că în cadrul ei există diferențe (inegalități) mai mari între elevi decît în cea dintîi. Desigur că și un asemenea indicator de împrăștiere este posibil; el există, poartă denumirea de amplitudine și se notează cu A . Expresia sa este diferența dintre valoarea maximă și cea minimă în cadrul seriei. Cu notațiile noastre:

$$A = x_s - x_1$$

Indicatorul amplitudine se poate aplica doar în situații foarte simple, cînd efectiv ne interesează numai reliefarea "spațiului de joc" al unei variabile. De cele mai multe ori, el este nesatisfăcător, deoarece face abstracție de frecvențe; într-adevăr problema dispersiei nu se reduce numai la intervalul valorilor luate, ci are în vedere și felul cum se împrăștie indivizii pe acest interval. Putem avea două caracteristici identice din punctul de vedere al valorilor, dar foarte diferite din cel al frecvențelor. Spre exemplificare, să recurgem, din nou, la două clase de elevi (imaginare) cărora li se dă un extemporal la limba română. Pentru a ușura citirea datelor, să presupunem că ele sînt de același efectiv; 4o elevi fiecare. Să zicem că distribuțiile notelor sînt cele din tabelul 1.

Note	4	5	6	7	8	9	1o	Total
Frecvențe:								
Clasa A	1	2	5	24	5	2	1	4o
Clasa B	4	5	6	1o	6	5	4	4o

Tabelul 1. Distribuțiile notelor în două clase, formate din cîte 4o elevi (cifre fictive)

Se observă că, în fiecare clasă, sînt luate toate notele de la 4 la 1o de către cel puțin un elev; deci amplitudinile sînt

aceleaşi: lo - 4 = 6 note. Mai mult, dacă privim cu atenţie, vom intui imediat că şi mediile notelor pe cele două clase sînt identice: nota 7 . Cu toate acestea, există o diferenţă importantă a modului cum se distribuie elevii noştri. In clasa A, media 7 este rezultatul unui număr foarte mare de note de 7 sau apropiate, pe cînd în B, deşi se menţine simetria în jurul lui 7, deşi această valoare este şi aici cea modală, totuşi, există un număr mult mai însemnat de indivizi, decît în clasa cealaltă, care se abat de la medie, în sus sau în jos. Putem deci afirma că cele două clase deşi se găsesc la acelaşi nivel mediu (median sau modal), diferă totuşi după gradul de dispersie. In prima avem de a face cu o omogenitate mai mare, în a doua întîlnim un grad superioar de dispersie sau de inegalitate: mai mulţi copii buni dar tot mai mulţi şi slabi.

Se pune deci problema calculării gradului de dispersie prin indicatori care să reflecte şi situaţii ca cea de mai sus. In principiu, există două posibilităţi de a surprinde dispersia indivizilor pe scala valorilor;

a) fără o valoare de referinţă faţă de care să se calculeze abaterile valorilor individuale şi

b) cu o asemenea valoare.

In primul caz, procedura de calcul este destul de complicată, căci trebuie luate în considerare toate cuplurile posibile de indivizi, pentru care se calculează diferenţa valorilor. Media acestor diferenţe luate în valoare absolută ne dă indicatorul de dispersie care poartă numele statisticianului italian Corrado Gini. Formula fiind mai complexă şi calculele mai laborioase, indicatorul lui Gini este rareori folosit, mai ales că el nu are avantaje de cunoaştere evitente faţă de alţi indici. In al doilea caz, este posibilă introducerea mai multor indicatori, diversitatea provenind din variantele multiple de alegere a valorii de referinţă şi de calcul al abaterilor faţă de ea. Voi enumera cîţiva, oprindu-mă puţin asupra celui mai important.

Abaterea medie faţă de o valoare a. Este, pur şi simplu, media distanţelor dintre valorile x_i şi punctul a, arbitrar ales, prin distanţe înţelegînd diferenţele x_i-a, în valoare absolută, multiplicată fiecare prin numărul de indivizi care iau valoarea x_i.

Formula va fi deci:

$$Am(a) = \frac{\sum_{i=1}^{s} k_i |x_i - a|}{n} \qquad (1)$$

<u>Abaterea medie față de medie</u> este un caz particular al indicatorului de mai sus, pentru situația cînd valoarea fixată, de referință, este chiar media. Se mai numește simplu <u>abatere medie</u> și se calculează cu formula:

$$Am = \frac{\sum_{i=1}^{s} k_i |x_i - \bar{x}|}{n} \qquad (2)$$

Aici se vede mai bine de ce este necesar ca diferențele să fie luate în valoarea absolută. Intr-adevăr, conform unei proprietăți a mediei, suma de la numărător fără valori absolute (suma abaterilor valorilor de la medie) este nulă. Așa însă, diferențele își pierd semnul și nu se mai anulează între ele, rezultatul fiind un număr pozitiv, la limită zero, dacă toate valorile coincid (între ele și deci și cu media lor). Cu cît distanțele valorilor față de medie sînt mai mari și cu cît aceste valori distanțate sînt luate de mai mulți indivizi, cu atît valoarea indicatorului va fi mai ridicată. Indicatorul va arăta, firește, așa cum sugerează și denumirea, cu cît se abate în medie un individ de la valoarea medie a caracteristicii.

După cum se știe, există și o altă modalitate de a înlătura semnul diferențelor decît modulul sau valoarea absolută. Ea constă în ridicarea acestora la pătrat. Procedura ridicării la pătrat este mai frecvent utilizată în statistică datorită algoritmului de calcul mai simplu. Desigur, mărimea astfel obținută este una pătratică și, pentru a se reveni la dimensiunea inițială a variabilei, trebuie extras radicalul. In cazul dispersiei, se vor obține două mărimi similare celor precedente.

<u>Abaterea medie pătratică față de o valoare</u> <u>a</u>. O vom nota cu S (a) sau, simplu, dacă nu trebuie precizată valoarea <u>a</u>, S. Formula de calcul va fi

$$S\ (a) = \sqrt{\frac{\sum_{i=1}^{s} k_i (x_i - a)^2}{n}} \qquad (3)$$

Făcînd a egal cu media, $\bar{x}$, se va obține indicatorul de dispersie cel mai cunoscut și mai utilizat în statistică: abaterea standard (σ):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{s} k_i (x_i - \bar{x})^2}{n}} \qquad (4)$$

In textele manualelor de statistică sau în alte lucrări din domeniul diferitelor științe, unde se utilizează indicatorul, el poate să mai apară și sub alte denumiri; deviație standard, ecart tip, sau pur și simplu, dispersie. Parțial, această varietate de denumiri în românește se datorează faptului că în engleză indicatorul se numește standard deviation iar în franceză, écart-type. Denumirea de dispersoe trebuie oricum înlăturată căci acesta este un termen generic pentru toată problematica indicatorilor respectivi, fiind sinonim cu cel de împrăștiere. Expresia franțuzească de astă dată nu este cea mai ușor de introdus în limba noastră căci termenul de ecart, deși cu oarecare uz, este totuși rareori întîlnit în texte. In expresia englezească, personal înclin - ca majoritatea, cred, a statisticienilor - spre înlocuirea cuvîntului deviație cu cel de abatere, care este consacrat în matematică și statistică pentru a desemna diferențe de genul celor ce intervin aici.

Pătratul abaterii standard (σ^2) - deci mărimea din formula (4) fără radical - are de asemenea, o largă folosire în statistică, motiv pentru care poartă un nume aparte: varianță. Indicatorul varianță intervine frecvent în analiza relațiilor dintre două sau mai multe variabile statistice. In aceste cazuri, varianța este considerată a fi expresia însăși de manifestare a unui fenomen sau, altfel spus, dar identic din punct de vedere statistic, ca o mărime ce reflectă gradul de nedeterminare a unei variabile. Explicarea unei variabile statistice prin altele este înțeleasă, în mod obișnuit, ca posibilitate de reducere a varianței variabilei

de explicat, atunci cînd variabilele explicative sînt menţinute constante sau, în alţi termeni, explicarea este reducerea gradului de nedeterminare în raport cu variabila dependentă, dacă se cunoaşte poziţia individului în raport cu celelalte variabile.

Să trecem acum la ilustrarea procedurilor de calculare a indicatorilor, în speţă a abaterii standard. Mai întîi să reluăm exemplul din tabelul 1, pentru a arăta, prin intermediul indicatorilor abatere medie şi abatere standard, diferenţa de dispersie dintre cele două clase de elevi. Vom aranja datele tabelului pe coloane pentru a facilita efectuarea produselor şi a însumărilor. Dat fiind că noi cunoaştem de la început media anbelor caracteristici - ea este 7 în cazul de faţă, calculul din schema ce urmează vizează exclusiv determinarea abaterii standard- De obicei, cum pentru determinarea mediei sînt necesare calcule de genul celor arătate mai înainte, schema de calcul de mai jos se poate lărgi, introducînd ca prime coloane, după cele ale valorilor şi frecgenţelor, pe cele cerute de calculul mediei.

Pentru clasa A:

x_i	k_i	$\lvert x_i-\bar{x}\rvert$	$(x_i-\bar{x})^2$	$k_i\lvert x_i-\bar{x}\rvert$	$k_i(x_i-\bar{x})^2$
4	1	3	9	3	9
5	2	2	4	4	8
6	5	1	1	5	5
7	24	0	0	0	0
8	5	1	1	5	5
9	2	2	4	4	8
1o	1	3	9	3	9
Total	4o			24	44

Deci

$$Am_A = \frac{24}{4o} = o,6o$$

şi

$$\sigma_A = \sqrt{\frac{44}{4o}} = \sqrt{1,1o} = 1,o5$$

unde prin indicele A s-a pus în evidenţă faptul că valorile calculate se referă la clasa şcolară (caracteristica) A.

Pentru clasa B:

x_i	k_i	$\lvert x_i - \bar{x} \rvert$	$(x_i - \bar{x})^2$	$k_i \lvert x_i - \bar{x} \rvert$	$k_i(x_i - \bar{x})^2$
4	4	3	9	12	36
5	5	2	4	10	20
6	6	1	1	6	6
7	10	0	0	0	0
8	6	1	1	6	6
9	5	2	4	10	20
10	4	3	9	12	36
Total	40			56	124

Si deci

$$Am_B = \frac{56}{40} = 1,40$$

$$\sigma_B = \sqrt{\frac{124}{40}} = \sqrt{3,10} = 1,76$$

Prin urmare, ambii indicatori - atît abaterea medie cît şi abaterea standard - pun în relief dispersia mai mare a notelor în clasa şcolară B, reflectînd într-o manieră cantitativă ceea ce s-a observat de la început în mod intuitiv. Cum era şi firesc, valorile celor doi indicatori, pentru o aceeaşi caracteristică, nu coincid, căci formulele de calcul sînt diferite. Abaterea medie arată efectiv că, de pildă în clasa B, notele elevilor se abat în medie de la media clasei cu 1,40 puncte. Pentru celălalt indicator, mai clară este interpretarea varianţei: la aceeaşi clasă B, media pătratelor abaterilor notelor individuale de la media clasei este de 3,10 puncte, iar radicalul acesteia este 1,76.

Să ne oprim, în continuare, doar la abaterea standard şi să prezentăm o metodă de simplificare a calculelor. In exemplele date, valorile medii au fost din întîmplare (mai exact datorită aranjării intenţionate a frecvenţelor pentru a ieşi aşa) cifre ro-

tunde şi deci nu a existat nici o dificultate de calcul. In general însă, mediile nu vor mai apare sub forme atît de simple şi deci vom ajunge la calcule mai laborioase atunci cînd facem diferenţele şi le ridicăm la pătrat. Pentru uşurarea muncii, ne vom folosi de o proprietate importantă a abaterii standard.

S-a văzut mai sus că abaterea standard reprezintă un caz particular de abatere medie pătratică (cînd a coincide cu $\bar{x}$). Fiind dat un a oarecare şi notînd:

$$d = a - \bar{x} \qquad (5)$$

atunci se poate scrie o relaţie simplă între $S^2(a)$ şi σ^2:

$$\sigma^2 = S^2 - d^2 \qquad (6)$$

Relaţia (6) arată că dintre toate abaterile medii pătratice, abaterea standard este cea mai mică. Intr-adevăr, dat fiind că d^2 este o mărime pozitivă sau nulă, din (6) rezultă imediat:

$$\sigma^2 \leq S^2$$

iar inegalitatea se transformă în egalitate atunci cînd d este nul, deci cînd a coincide cu media.

Formula (6) poate fi folosită pentru a reduce volumul calculelor, deoarece ea ne arată că, înainte de calcularea lui σ^2, putem determina un S^2, cu o valoare a convenabil aleasă în acest scop, urmînd ca media să intervină în lucru numai o dată, la sfîrşit, în calculul lui d^2. Pentru ilustrare, să reluăm exemplul cu dimensiunea gospodăriilor din localitatea rurală. In paragraful precedent am calculat valoarea medie:

$$\bar{x} = 3,66$$

şi vom continua acum cu determinarea abaterii standard. Să luăm ca valoare ajutătoare:

$$a = 4$$

Va rezulta

$$d = 4 - 3,66 = 0,34$$

şi

$$d^2 = 0{,}1156$$

Calculele ce urmează ne duc la determinarea lui $S^2(4)$:

x_i	k_i	$x_i - a$	$(x_i - a)^2$	$k_i(x_i - a)^2$
1	3o	-3	9	27o
2	4o	-2	4	16o
3	6o	-1	1	6o
4	5o	0	0	0
5	3o	1	1	3o
6	2o	2	4	8o
7	1o	3	9	9o
8	5	4	16	8o
9	5	5	25	125
Total	25o			895

$$S^2 = \frac{895}{250} = 3{,}58$$

$$\sigma^2 = 3{,}58 - 0{,}1156 = 3{,}4644$$

$$\sigma = 1{,}86$$

Observaţie: După cum se poate constata uşor, abaterea standard, ca şi ceilalţi indicatori introduşi aici, se poate scrie şi cu frecvenţele relative. Intr-adevăr, împărţind fiecare termen al sumei cu n, formula (4) devine:

$$\sigma = \sqrt{\sum_{i=1}^{s} p_i(x_i - \bar{x})^2}$$

Din formulele de calcul rezultă că toţi coeficienţii de dispersie sînt mărimi concrete, în sensul că posedă aceeaşi unitate

de măsură ca şi variabila pe care o reprezintă. Acest lucru facilitează, desigur, înţelegerea semnificaţiei lor dar, în acelaşi timp reprezintă şi un impediment pentru comparaţii. Dacă pentru indicatorii de poziţie - exprimaţi şi ei în aceleaşi unităţi - problema comparaţiei nu se pune decît pentru variabile identice definite pe populaţii diferite (nu are sens să comparăm, de pildă,vîrsta medie cu media numărului de zile lucrate în CAP, ci doar vîrsta medie a cooperatorilor dintr-un sat cu cea a cooperatorilor din altul), pentru cei de împrăştiere există situaţii în care comparaţiile apar necesare şi pentru variabile diferite definite pe o aceeaşi populaţie şi chiar - e drept mai rar - pentru variabile diferite rezultate din studiul a două comunităţi distincte. De exemplu, în cadrul unei colectivităţi ne putem pune problema dacă inegalitatea indivizilor în raport cu veniturile obţinute este mai mare sau mai mică decît în raport cu o altă variabilă cum ar fi consumul cultural (presupunînd că s-a putut măsura acesta).

Este deci clar că utilizarea în comparaţii a abaterii standard este posibilă doar cînd variabilele sînt aceleaşi, cum a fost cazul cu notele la extemporale în cele două clase A şi B, In alte situaţii, de genul celei imaginate mai sus, va trebui să se calculeze indicatori de dispersie care să nu depindă de unitatea de măsură. Unul dintre aceştia este aşa-numitul <u>coeficient de variaţie</u> v, dat de formula:

$$v = \frac{\sigma}{\bar{x}}$$

sau, sub formă procentuală:

$$v = \frac{\sigma}{\bar{x}} \cdot 100$$

Prin împărţirea cu media, dispare unitatea de măsură şi indicatorul exprimă proporţia sau procentul reprezentat de abaterea standard din medie.

Unele manuale de statistică indică faptul că acest coeficient este expresiv doar atunci cînd media nu se află în apropierea lui zero, căci altfel valoarea lui este foarte ridicată. Din cele expuse în capitolul referitor la scalele de măsură, se va înţelege de noi că un astfel de indicator se poate utiliza, strict vorbind, doar cînd scala este de rapoarte, deci cînd avem de a face cu un zero natural. Intr-adevăr, în cazul scalei de interval, valoarea coeficientului este sensibilă la orice translaţie a scalei căci media se modifică în mod corespunzător, pe cînd abaterea standar rămîne nemodificată. In particular, orice scală de intervale poate fi schimbată astfel ca media să fie nulă şi atunci coeficientul de variaţie nu poate fi calculat.

Importanţa indicatorilor de dispersie pentru analiza statistică se poate cu greu intui în întregime după cele spuse mai sus. Abia în capitolele următoare se va întrezări ceva mai clar locul privilegiat pe care aceştia îi ocupă în analizele statistice, îndeosebi în încercările de a modela relaţiile dintre fenomene.

Cap. IX. A S O C I E R E A

1. Cazul caracteristicilor dihotomice

Studiul realizat pînă în prezent asupra caracteristicilor sau variabilelor statistice, avînd în vedere tot timpul doar un singur factor în raport cu care se efectuează distribuţia indivizilor în clasele (sau după valorile) caracteristicii (variabilei), poate fi considerat, în esenţă, doar o descriere a populaţiei în funcţie de factorul respectiv. Urmează în continuare să încercăm a realiza cîţiva paşi pe drumul ce duce la explicarea fenomenelor sociale, utilizînd în acest scop mijloace statistice adecvate.

Orice raţionament (iar explicaţia nu este decît un lanţ de raţionamente) porneşte de la anumite propoziţii (judecăţi) iar orice judecată presupune două noţiuni care se "leagă" una de cealaltă. Tradus în limbajul nostru - şi o să vedem imediat cum se face această traducere - asta înseamnă o analiză a două caracteristici statistice. Dar nu este vorba de a realiza o descriere separată a fiecăreia din ele, ci de o analiză din care să rezulte comportamentul indivizilor simultan în raport cu cele două caracteristici.

Pentru a înţelege mai uşor ce înseamnă această legătură vom începe cu cazul cel mai simplu şi anume cel al clasificării dihotomice. O clasificare de acest tip corespunde schemei clasice de analiză a unei noţiuni după sfera şi conţinutul său. Mai exact, o noţiune definită pe o mulţime de n indivizi împarte populaţia noastră în două submulţimi: A (sau A_1)[x] - cei care posedă trăsătura definitorie - şi $\bar{A}$ (sau A_2) - cei care nu o posedă. Exemplu: fumători/nefumători, religioşi/atei etc. Aşa cum ştim, o asemenea caracteristică se poate reda printr-un tabel de forma:

$A(A_1)$	$\bar{A}(A_2)$	Total
k_1	k_2	n

x - Voi folosi pentru clase fie notaţia A şi $\bar{A}$, fie A_1 şi A_2, de la caz la caz.

unde k_1 şi k_2 ne arată <u>cîţi</u> indivizi din cei <u>n</u> sînt A (sau A_1) şi cîţi sînt $\bar{A}$ (sau A_2).

In mod analog, după un alt atribut B se realizează o altă clasificare de acelaşi tip: o parte dintre indivizi vor intra în clasa B (sau B_1) iar alta în $\bar{B}$ (sau B_2). Dar cum, de regulă, cele două clasificări nu coincid, este clar că în final se vor obţine 4 submulţimi în populaţia celor <u>n</u> indivizi:

- A şi B (A_1 şi B_1), cuprinzînd k_{11} indivizi
- A şi $\bar{B}$ (A_1 şi B_2) , " k_{12} "
- $\bar{A}$ şi B (A_2 şi B_1) " k_{21} "
- $\bar{A}$ şi $\bar{B}$ (A_2 şi B_2) " k_{22} "

Prin urmare, din totalul de <u>n</u> indivizi, k_{11} se găsesc simultan în A şi B (sau în A_1 şi B_1), k_{12} se află în A şi $\bar{B}$ etc. Această diviziune se poate pune sub forma unui tabel ca cel de mai jos:

A \ B	B(B_1)	$\bar{B}$(B_2)	Total
A(A_1)	k_{11}	k_{12}	$k_{1.}$
$\bar{A}$(A_2)	k_{21}	k_{22}	$k_{2.}$
Total	$k_{.1}$	$k_{.2}$	n

Tabelul 1.
<u>Notaţii</u>

unde semnificaţia frecvenţelor cu doi indici este cea explicată mai sus. Cît priveşte frecvenţele cu un indice şi un punct, care poartă denumirea de <u>frecvenţe marginale</u> (cele cu doi indici fiind numite <u>interioare</u>), ele nu redau altceva decît distribuţiile indivizilor după fiecare caracteristică separat. Rezultă imediat relaţiile:

$$k_{11} + k_{12} = k_{1.} \qquad k_{21} + k_{22} = k_{2.}$$

$$k_{11} + k_{21} = k_{.1} \qquad k_{12} + k_{22} = k_{.2}$$

$$k_{1.} + k_{2.} = k_{.1} + k_{.2} = k_{11} + k_{12} + k_{21} + k_{22} = n$$

Tabelul 1 poartă denumirea de <u>tabel de asociere</u> a caracteristicilor A şi B. Se observă din relaţiile scrise mai sus că dacă frecvenţele marginale sînt date, atunci este posibilă fixarea unei singure frecvenţe interioare, celelalte trei rezultă ca diferenţe la totalurile de pe liniile sau coloanele tabelului. Exemplu:

$$k_{12} = k_{1.} - k_{11} \qquad \text{etc.}$$

Din acest motiv se spune că tabelul 1 are un singur <u>grad de libertate.</u>

Pentru o mai bună înţelegere a celor expuse pînă aici să dăm un exemplu concret. Să presupunem că într-un studiu asupra unei populaţii de 1oo indivizi am stabilit două caracteristici: A - sexul (cu clasele A_1 - bărbaţi şi A_2 - femei) şi B - comportamentul în privinţa fumatului (cu clasele B_1 - fumători şi B_2 - nefumători). Rezultatul - ipotetic - obţinut este redat în tabelul 2.

A \ B	Fumători	Nefumători	Total
Bărb.	3o	2o	5o
Fem.	2o	3o	5o
Total	5o	5o	1oo

Tabelul 2.

Avem deci 3o bărbaţi fumători, 2o bărbaţi nefumători, 2o femei fumătoare şi 3o femei nefumătoare. In total, pe fiecare caracteristică, avem: 5o bărbaţi şi 5o femei; 5o fumători şi 5o nefumători. Cu notaţiile de mai sus:

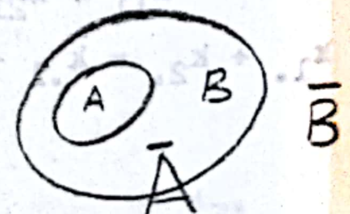
$$k_{11} = 3o \,, \quad k_{12} = 2o; \quad k_{1.} = 5o \qquad \text{etc.}$$

Pentru a facilita înţelegerea modului de interpretare a unor tabele ca cel de mai sus, voi face apel la una dintre relaţiile logice foarte cunoscute: <u>implicaţia</u> . O implicaţie de tipul A   B poate fi formulată astfel:

" dacă A atunci B"

sau

"toţi A sînt B"



cu reprezentarea schematică a sferelor noţiunilor ca mai sus. Consecinţa unei astfel de implicaţii, consecinţă cu repercusiuni asupra formei tabelului nostru, va fi aceea că nu va exista nici un individ care să fie în acelaşi timp A şi $\bar{B}$, prin urmare una dintre căsuţele interioare ale tabelului va conţine o frecvenţă nulă:

$$k_{12} = 0$$

De aici decurge automat că vom avea:

$$k_{11} = k_{1.} \text{ şi } k_{22} = k_{.2}$$

Tabelul va fi deci de forma:

B \ A	B_1	B_2	Total
A_1	k_{11}	0	$k_{1.}$
A_2	k_{21}	k_{22}	$k_{2.}$
Total	$k_{.1}$	$k_{.2}$	n

Tabelul 3.

Forma tabelului ce corespunde implicaţiei $A \longrightarrow B$.

Un tabel de formă asemănătoare se va obţine în cazul în care implicaţia merge de la B spre A. Atunci, pentru că toţi B sînt A, nu vom avea nici un individ în intersecţia lui B cu $\bar{A}$ şi deci se va anula frecvenţa k_{21}. Vor decurge egalităţile:

$$k_{11} = k_{.1} \text{ şi } k_{22} = k_{2.}$$

In fine, dacă implicaţia este reciprocă ($A \longleftrightarrow B$), deci toţi A sînt B şi toţi B sînt A, ambele frecvenţe de pe diagonala secundară se vor anula:

$$k_{12} = k_{21} = 0$$

şi prin urmare:

$$k_{11} = k_{1.} = k_{.1} \text{ ; } k_{22} = k_{2.} = k_{.2}$$

Rezultă deci că <u>o implicaţie reciprocă strictă se poate realiza numai atunci cînd structurile determinate de cele două caracteristici în populaţie coincid</u>. In cazul în care mai puţini indivizi posedă atributul A decît B atunci putem avea cel mult o implicaţie univocă de la A spre B şi invers.

Trebuie spus însă că asemenea situaţii de implicaţii stricte nu se întîlnesc prea des în realitatea socială şi chiar dacă se întîlnesc nu reflectă decît legături banale sau evidente între fenomene. Aceasta este încă o dovadă că faptele sociale se găsesc, aşa cum am mai spus, în relaţii de determinare statistică şi nu dinamică. Prin urmare interesul raporturilor descrise mai sus rezidă doar în faptul că ele sînt nişte cazuri ideale, faţă de care realitatea se îndepărtează mai mult sau mai puţin. Dacă dorim totuşi să dăm un exemplu concret de implicaţie strictă am putea imagina situaţia următoare. Să presupunem că într-o cercetare concretă am suprins comportamentul unui lot de 1oo indivizi în raport cu frecventarea bisericii (caracteristica A) şi pe de altă parte am reuşit să-i delimităm pe cei care cred în religie de cei care nu cred (B). Rezultatul ar putea apare ca în tabelul 4.

Tabelul 4

A \ B	Credincioşi	Necred.	Total
Merg regulat la biserică	3o	0	3o
Nu merg regulat la biserică	1o	6o	7o
	4o	6o	1oo

Avem deci în cazul tabelului 4 o ipotetică situaţie de implicaţie strictă de la A spre B: "Toţi cei care merg regulat la biserică sînt credincioşi", deşi poate că în realitate nici în astfel de cazuri nu obţinem frecvenţa 0 (s-ar putea ca, dintr-un motiv sau altul, în perioada cît efectuăm observaţua vreun necredincios să frecventeze cu regulatitate biserica).

Situaţiile de implicaţie strictă pot fi interpretate şi în altă terminologie. In limbajul teoriei informaţiei, putem apre-

cia că pentru un individ care se află în clasa A (frecventează regulat biserica) deținem o informație completă asupra felului în care se va plasa în clasele lui B: el se plasează cu necesitate în B și nu în B̄. Sau în limbajul teoriei probabilităților: probabilitatea de a fi B a unui individ care este A este egală cu 1.

Ar mai fi de făcut o precizare referitoare la funcționalitatea relațiilor de implicație strictă în cadrul explicației, mai exact la raportul acestora cu relațiile cauzale. Fără a intra în probleme de epistemologie, precizez doar că, după opinia mea, implicația nu poate fi mai mult decît un indicator care să sugereze o eventuală legătură cauzală. Nu numai că ea nu poate să garanteze o asemenea legătură, dar nici nu-i poate măcar sugera sensul. Iată, chiar în cazul exemplului din tabelul 4, faptul că A ⟶ B nu înseamnă că avem o determinare cauzală în același sens. Nu faptul că oamenii merg mai des la biserică determină gradul lor de religiozitate, ci mai degrabă invers. In general, atitudinile și opiniile oamenilor le determină comportamentul și nu invers, deși, evident, între acești doi factori pot exista, cum probabil se întîmplă și în cazul exemplului nostru, interdependențe, reacții de influențare reciprocă.

Concluzia observației precedente este una de o importanță covîrșitoare pentru toți cei care utilizează instrumentarul statistic în analiza fenomenelor sociale: numai dintr-o relație statistică, oricît de perfectă ar fi ea, nu se pot trage concluzii asupra fenomenelor reale studiate. Numai introducînd ceva din afară, mai bine zis din domeniul teoriei sau al schemelor noastre explicative, constituite înaintea cercetării, de care sîntem sau nu conștienți putem interpreta o legătură statistică.

Dar să mergem mai departe cu analiza. Am spus că în realitate cazurile pure în care sfera unei noțiuni este cuprinsă în sfera alteia - sau, mai mult, cînd ele coincid - sînt rare și neinteresante. Ca urmare ne vor preocupa situațiile în care sferele celor două noțiuni se intersectează. Aceasta logic ar suna: "Unii A sînt B". Statistic însă ne interesează să vedem exact cîți din A sînt și în B, mai bine zis, în ce proporție cei care sînt A sînt și B, în raport cu cei care se găsesc în A. Am putea continua demersul urmînd o cale propusă de R.Boudon prin introducerea noțiunii de implicație slabă, în contrast cu implicațiile stricte utilizate mai sus. O

implicaţie slabă de la A spre B se poate formula: "Dacă A ,atunci mai degrabă B decît B̄". Tot aşa se poate da o expresie şi implicaţiei slabe de la B spre A precum şi celei reciproce. Un exemplu de implicaţie slabă de la A spre B este dat în tabelul 5.

A \ B	B	B̄	Total
A	3o	lo	4o
Ā	3o	3o	6o
Total	6o	4o	loo

Tabelul 5.

Se vede că "dacă A, atunci mai degrabă (mai probabil) B decît B̄" , căci din cei 4o de indivizi din A, 3o se plasează în B şi numai lo în B̄. Din contră, dacă B, atunci în egală măsură A şi Ā (3o la 3o).

In realitate, şi cazuri ca acesta sînt rare, în sensul că este puţin probabil ca indivizii dintr-o clasă a unei caracteristici să se găsească exact cîte 5o% în clasele celeilalte. In consecinţă, de cele mai multe ori vom întîlni implicaţii slabe reciproce, pentru analiza cărora se raţionează în felul următor.

Dintre cele două caracteristici alegem una care ar putea determina (cauza, condiţiona etc.) pe cealaltă. In cazul în care o asemenea alegere este imposibilă alegem drept bază de plecare oricare din cele două caracteristici. Fie cea aleasă A. Vom încerca apoi să vedem cum se comportă indivizii populaţiei noastre vis-à-vis de cealaltă caracteristică B, în funcţie de apartenenţa lor la una sau alta din clasele lui A. Mai precis, vom calcula proporţia indivizilor din A care sînt în B, pe care o vom compara cu cea a indivizilor din Ā care sînt în B. De exemplu, în cazul tabelului 2, caracteristica de bază este evident sexul, care poate fi un factor ce influenţează fumatul. Atunci vom spune:

- dintre bărbaţi (clasa A sau A_1 - după notaţiile noastre) 3o din 5o sînt fumători;
- dintre femei (Ā sau A_2) 2o din 5o fumează;
- în consecinţă, dintre bărbaţi fumează mai mulţi (sau fumatul este mai intens practicat de bărbaţi) deoarece raportul 3o/5o

este mai mare decît 2o/5o.

In plus, dacă vrem să acordăm o semnificaţie mai precisă expresiilor "mai mulţi " sau "mai intens" este normal că cel mai simplu lucru este să calculăm diferenţa celor două rapoarte:

$$\frac{3o}{5o} - \frac{2o}{5o} = \frac{1o}{5o} = o{,}2$$

diferenţă care ne dă tocmai o măsură a diferenţei de comportament a celor două sexe în privinţa fumatului.

Poate mai uşor de înţeles stau lucrurile dacă în loc de rapoarte vom lucra cu procente. Calculăm, în cazul de faţă, procentele de fumători şi nefumători în rîndul bărbaţilor, apoi la femeilor şi apoi în totalul populaţiei. Se obţine tabelul 6.

A \ B	Fumăt.	Nefumăt.	Total
Bărb.	6o	4o	1oo
Fem.	4o	6o	1oo
Total	5o	5o	1oo

Tabelul 6. Procentele de fumători şi nefumători în rîndul bărbaţilor, femeilor şi a populaţiei totale studiate, corespunzător tabelului 2.

Se vede acum că dintre bărbaţii anchetaţi 60% fumează, dintre femei 40% iar pe ansamblu procentul de fumători este de 50%. Prin urmare, diferenţa dintre bărbaţi şi femei este de 20%, corespunzînd diferenţei de rapoarte anterioare de o,2.

Generalizînd modalitatea de analiză de mai sus prin folosirea notaţiei din tabelul 1, vom calcula diferenţa de comportament a indivizilor din cele două clase ale lui A în raport cu caracteristica B astfel:

$$\delta_A = \frac{k_{11}}{k_{1.}} - \frac{k_{21}}{k_{2..}}$$

In cazul particular al tabelului 2, valoarea lui δ_A va fi tocmai o,2.

In mod analog, se poate calcula un indicator simetric, care să reflecte diferenţa de comportament a indivizilor din cele două clase alelului B vis-à-vis de caracteristica A:

$$\delta_B = \frac{k_{11}}{k_{.1}} - \frac{k_{12}}{k_{.2}}$$

In cazul tabelului 2, vom avea:

$$\delta_B = \frac{3o}{5o} - \frac{2o}{5o} = o,2$$

ceea ce înseamnă că cei doi indicatori coincid. Desigur, această coincidenţă nu este obligatorie, ci ea se produce doar în anumite situaţii particulare. Se poate demonstra că cei doi indicatori de diferenţă, δ_A şi δ_B, coincid atunci şi numai atunci cînd se realizează cel puţin una din următoarele trei condiţii:

a) $k_{1.} = k_{.1}$

b) $k_{1.} + k_{.1} = n$

c) $k_{11} = k_{1.} k_{.1}/n$

Condiţia a) înseamnă coincidenţa structurilor determinate în populaţie de cele două caracteristici, căci ea implică şi $k_{2.} = k_{.2}$. A doua condiţie înseamnă tot coincidenţa celor două structuri numai că în sens invers; din b) rezultă $k_{1.} = k_{.2}$ şi $k_{.1} = k_{2.}$. Despre cea de a treia vom vorbi imediat.

Tot despre cele două valori δ_A şi δ_B se pot demonstra următoarele două proprietăţi fundamentale:

i) Cele două diferenţe nu pot fi simultan una de un semn şi alta de semn contrar.

ii) Cele două diferenţe nu pot fi simultan una egală cu şi alta diferită de zero.

Pomenind de semnul indicatorilor δ am admis deci că ei pot lua şi valori negative. Să analizăm un astfel de caz (tab.7) în care cele două caracteristici sînt: A - rata brută de natalitate,

B - PNB pe cap de locuitor, populaţia statistică fiind constituită din 125 de ţări ale lumii.

A \ B	PNB mare	PNB mic	Total
Natal. mare	9	88	97
Natal. mică	26	2	28
Total	35	90	125

Tabelul 7.

De această dată, B apare ca un determinant mai probabil al lui A decît invers. Aşadar, în primă instanţă, vom calcula

$$\delta_B = \frac{9}{35} - \frac{88}{90} = 0,257 - 0,978 = - 0,721$$

care se poate interpreta astfel:

- Ţările cu PNB mare au natalitate mare în proporţie de 27,7%
- Ţările cu PNB mic au natalitate mare în proporţie de 97,8%
- Deci avem o diferenţă de 72,1% de natalitate mare în favoarea ţărilor sărace.

Calculînd şi indiactorul simetric δ_A, se obţine:

$$\delta_A = \frac{9}{97} - \frac{26}{28} = 0,093 - 0,929 = - 0,835$$

deci tot o valoare negativă, relativ apropiată de prima dar totuşi diferită, dată fiind necoincidenţa celor două structuri marginale.

In sfîrşit, putem trece acum să precizăm şi să definim noţiunea de <u>asociere</u>. Se observă că, dacă analizăm simultan comportamentul indivizilor în raport cu cele două caracteristici, plecînd de la o clasă a uneia - să zicem A -, atunci ei au tendinţa de a se grupa în măsură mai mare într-o clasă a lui B, decît o fac cei care se găsesc în acelaşi timp în A. Si invers, plecînd de la B spre A, găsim asemenea tendinţe. In general, există două mari posibilităţi de astfel de grupări: fie cei care sînt în A se plasea-

ză preponderent în B (şi atunci automat cei din $\bar{A}$ se grupează în $\bar{B}$) , fie cei din A se plasează de predilecţie în $\bar{B}$ (şi atunci cei din $\bar{A}$ vor avea tendinţa să se "îngrămădească" în clasa B). In primul caz vom avea o concentrare pe diagonala principală a tabelului (frecvenţele k_{11} şi k_{22}) iar în al doilea pe diagonala secundară, corespunzător frecvenţelor k_{12} şi k_{21}.

Această tendinţă a indivizilor de a nu se distribui uniform în tabel şi de a se concentra pe una sau alta din cele două diagonale reprezintă conţinutul noţiunii de asociere. De aici şi denumirea care se dă în general tabelelor de tipul 1 - tabele de asociere.

Cu ajutorul consideraţiilor anterioare, putem defini în mod riguros noţiunea de asociere.

Vom spune că într-un tabel de tipul tabelului 1 există o asociere între caracteristicile A şi B dacă valorile δ sînt diferite de zero. Vom spune că cele două caracteristici A şi B sînt neasociate sau independente dacă valorile δ se anulează. (Am menţionat că cele două valori δ nu pot fi simultan una zero şi alta nenulă). Dacă cele două valori δ au ambele semnul plus, atunci asocierea se spune că este pozitivă, situaţie care corespunde concentrării indivizilor pe diagonala principală a tabelului. In caz contrar asocierea este negativă. (Am precizat, de asemenea, că cele două valori δ , dacă sînt nenule, trebuie să aibă acelaşi semn.)

Referitor la noţiunea de semn sau sens al asocierii se cere făcută o precizare. De multe ori semnul asocierii apare ca ceva artificial, în funcţie de felul cum se aşează clasele caracteristicilor în tabel. Aşa de exemplu, în cazul tabelului 2 vom avea o asociere pozitivă (în sensul precizat că în cei doi δ sînt pozitivi - o,2 - sau că indivizii se concentrează pe diagonala principală). Dacă însă vom scrie întîi linia femeilor şi apoi a bărbaţilor, semnul asocierii se va inversa şi el. Ca urmare în astfel de cazuri se cere precizat ce se înţelege prin asociere pozitivă sau negativă: vom zice că există o asociere între sex şi fumat, în sensul că bărbaţii fumează în proporţie mai mare decît femeile. Nu acelaşi lucru se întîmplă în cazul tabelului 7. Aici este suficient să spunem că între natalitate şi produsul naţional brut pe cap de locuitor există o asociere negativă ca să se înţeleagă faptul că ţările cu PNB mare au natalitate mică şi invers. In conclu-

zie, putem prezenta ca regulă următoarea: dacă, fixînd clasele unei caracteristici, urmează automat o ordine şi pentru clasele celelalte, atunci are sens să vorbim de un semn al asocierii; în caz contrar, nu şi trebuie să explicăm direcţia asocierii.

După depistarea existenţei (inexistenţei) şi a semnului asocierii (în cazul că există), urmează al treilea pas, care va consta în încercarea de a măsura intensitatea asocierii, respectiv de a construi un indice care să reflecte gradul în care cele două caracteristici sînt legate una de cealaltă.

Există de fapt mai mulţi asemenea indici propuşi în literatura de specialitate statistică. Unul dintre ei, foarte răspîndit, pe care-l vom numi pur şi simplu coeficient de asociere se poate calcula pornind tocmai de la cele două valori δ . Coeficientul de asociere, notat cu litera γ , se poate obţine luînd media geometrică a lui δ_A şi δ_B, ataşîndu-i semnul + sau - în concordanţă cu semnul diferenţelor δ . Deci:

$$\gamma = \pm\sqrt{\delta_A \delta_B} \qquad (1)$$

Se poate demonstra uşor că la aceeaşi valoare se poate ajunge direct din frecvenţele tabelului 1. Cu respectivele notaţii:

$$\gamma = \frac{nk_{11} - k_{1.}k_{.1}}{\sqrt{k_{1.}k_{2.}k_{.1}k_{.2}}} \qquad (2)$$

formulă în care nu se mai cere atribuirea explicită a unui semn, deoarece acesta rezultă imediat din semnul diferenţei de la numărător.

Exemple:

1. Să aplicăm formula (2) tabelului 2.

$$\gamma = \frac{100x300 - 50x50}{\sqrt{50x50x50x50}} = \frac{3000 - 2500}{2500} = \frac{500}{2500} = 0,2$$

Intr-adevăr, aici trebuia să se obţină valoarea 0,2 căci, după cum s-a văzut, atît δ_A cît şi δ_B aveau aceeaşi valoarea 0,2, deci

şi media lor geometrică trebuia să iasă tot atît.

2. Pentru situaţia din tabelul 4:

$$\delta_A = \frac{30}{30} - \frac{10}{70} = 1 - 0,14 = 0,86$$

$$\delta_B = \frac{30}{40} - \frac{0}{60} = 0,75 - 0 = 0,75$$

şi

$$\varphi = \sqrt{0,86 \times 0,75} = 0,80$$

sau după forma (2):

$$\varphi = \frac{100 \times 30 - 30 \times 40}{\sqrt{30 \times 70 \times 40 \times 60}} = \frac{3000 - 1200}{100\sqrt{504}} = \frac{1800}{2245} = 0,80$$

3. Pentru tabelul 7:

$$\varphi = - \sqrt{0,721 \times 0,835} = - 0,776$$

sau

$$\varphi = \frac{9 \times 125 - 97 \times 35}{\sqrt{97 \times 28 \times 35 \times 90}} = - \frac{2270}{2925} = - 0,776$$

Proprietăţi ale coeficientului de asociere:

i) Limitele de variaţie a coeficientului φ sînt de la -1 la +1.

ii) Valorile -1 şi +1 sînt luate atunci şi numai atunci cînd toţi indivizii se plasează pe diagonala principală, respectiv, secundară a tabelului, deci în cazurile în care avem implicaţiile stricte reciproce $A \longleftrightarrow B$ sau $A \longleftrightarrow \bar{B}$. Am văzut că o condiţie necesară pentru a se obţine asemenea implicaţii o constituie coincidenţa structurilor determinate de A şi B în populaţie, căci toţi A trebuie să fie B şi reciproc (sau toţi A trebuie să fie $\bar{B}$ şi reciproc). Prin urmare, în cazul implicaţiilor stricte univoce (numai de la A spre B sau numai de la B spre A), valoarea lui φ nu poate fi 1 sau -1; este cazul situaţiei din tabelul 4, unde $\varphi = 0,8$.

iii) Valoarea 0 va fi luată atunci cînd diferențele sînt nule, deci în cazul de neasociere sau independență a caracteristicilor. Din formula (2) rezultă o condiție necesară și suficientă pentru independență prin anularea numărătorului fracției:

$$nk_{11} - k_{1.}k_{.1} = 0$$

sau

$$k_{11} = \frac{k_{1.}k_{.1}}{n} \qquad (3)$$

Se observă că acesta constituie cazul al treilea, c), în care cele două diferențe δ_A și δ_B coincid (vezi p. 228) Legătura dintre cele trei valori fundamentale ale lui φ (-1, 0, +1) și coincidența lui δ_A cu δ_B este ușor de înțeles. Datorită faptului că φ este media geometrică a celor două diferențe, el nu poate fi egal cu 1 decît dacă acestea din urmă coincid fiecare cu 1. De asemenea, φ se anulează atunci cînd diferențele δ se anulează. Deci o condiție necesară, dar evident nu suficientă, ca φ să ia valorile critice +1,0 și -1 este ca valorile δ să coincidă.

Mai prezint două noțiuni, legate de fapt de proprietatea ii). Un tabel care redă o situație de implicație strictă univocă, într-un sens sau altul, vom spune că este un tabel de asociere totală sau maximă, iar unul care prezintă o implicație strictă reciprocă se va numi de asociere perfectă.

Altă metodă de calcul a coeficientului de asociere

Pornind de la definiția asocierii sau independenței putem raționa și în felul următor. Avînd dat tabelul 1, ca rezultat al unei investigații empirice, ne punem problema de a vedea dacă și în ce măsură frecvențele sale interioare diferă de cele ale tabelului care, avînd aceleași distribuții marginale, nu prezintă nici o asociere. Frecvențele unui asemenea tabel, pe care-l numim tabel de independență frecvențe pe care le vom nota cu m_{ij}, se vor obține cu ajutorul formulei (3):

$$m_{ij} = \frac{k_{i.}k_{.j}}{n} \qquad i,j = 1,2 \qquad (3')$$

Dacă toate frecvenţele k_{ij} coincid cu frecvenţele teoretice corespunzătoare m_{ij}, atunci ne vom găsi în cazul de independenţă, dacă nu, atunci vom avea o asociere în tabel. Diferenţe pot exista numai în două sensuri:

- fie $k_{11} > m_{11}$, de unde rezultă automat inegalităţile: $k_{22} > m_{22}$ şi $k_{12} < m_{12}$ şi $k_{21} < m_{21}$, deci indivizii se concentrează pe diagonala principală a tabelului,

- fie $k_{11} < m_{11}$, de unde rezultă inegalităţi inverse celor de mai sus, deci indivizii se concentrează pe diagonala secundară.

In felul acesta, prin comparaţia valorilor k şi m se poate depista simultan atît existanţa cît şi sensul asocierii. Rămîne problema măsurării ei. Pentru aceasta se face apel la o celebră mărime statistică, notată cu χ^2 (hi pătrat), dată, în cazul nostru, de formula:

$$\chi^2 = \frac{(k_{11}-m_{11})}{m_{11}} + \frac{(k_{12}-m_{12})^2}{m_{12}} + \frac{(k_{21}-m_{21})^2}{m_{21}} + \frac{(k_{22}-m_{22})^2}{m_{22}}$$

sau, mai concentrat

$$\chi^2 = \sum_{i=1}^{2}\sum_{k=1}^{2} \frac{(k_{ij}-m_{ij})^2}{m_{ij}} \qquad (4)$$

Făcînd diferenţele căsuţă cu căsuţă, mărimea χ^2 măsoară evident abaterea dintre tabelul empiric şi cel cu frecvenţe întîmplătoare, adică tocmai gradul de asociere din tabel. Dar pentru că χ^2 depinde de n, numărul indivizilor din populaţie, şi nu se încadrează într-un interval de variaţie precis, nu poate fi utilizată ca indice de asociere. Se demonstrează fără dificultate că se poate trece uşor de la χ^2 la φ prin formula de mai jos:

$$\varphi = \pm\sqrt{\frac{\chi^2}{n}} \qquad (5)$$

semnul atribuindu-se în funcţie de situaţiile descrise mai înainte.

Exemple

1. Să presupunem că în urma unei cercetări s-au construit două caracteristici A şi B - a căror semnificaţie nu ne interesează aici - şi că asocierea dintre ele se realizează ca în tabelul 8. In tabelul 9 apar frecvenţele ce corespund ipotezei de independenţă.

A \ B	B	B	Total
A	1o	4o	5o
$\bar{A}$	3o	2o	5o
Total	4o	6o	1oo

Tabelul 8. Frecvenţe empirice

A \ B	B	B	Total
A	2o	3o	5o
$\bar{A}$	2o	3o	5o
Total	4o	6o	1oo

Tabelul 9. Frecvenţe teoretice

Calculul frecvenţelor m_{ij} din tabelul 9 se realizează după formula (3'):

$$m_{11} = \frac{5o\text{x}4o}{1oo} = 2o\ ; \qquad m_{12} = \frac{5o\text{x}6o}{1oo} = 3o \text{ etc.}$$

Vom avea:

$$\chi^2 = \frac{(1o-2o)^2}{2o} + \frac{(4o-3o)^2}{3o} + \frac{(3o-2o)^2}{2o} + \frac{(2o-3o)^2}{3o} =$$

$$= \frac{1oo}{2o}\ \frac{1oo}{3o}\ \frac{1oo}{2o}\ \frac{1oo}{3o} = 5 + 3;\ 3 + 5 + 3,3 = 16,6$$

Si

$$\varphi = -\sqrt{\frac{16,6}{1oo}} = -\ o,41$$

Semnul - a fost acordat în urma observaţiei că în tabelul 8 frecvenţele empirice se concentrează pe diagonala secundară.

2. Să reluăm exemplul din tabelul 2. In situaţia de independenţă între caracteristicile sex şi fumat, s-ar fi obţinut:

A \ B	Fumători	Nefumăt.	Total
Bărbaţi	25	25	5o
Femei	25	25	5o
Total	5o	5o	1oo

Si deci

$$\chi^2 = \frac{(3o-25)^2}{25} + \frac{(2o-25)^2}{25} + \frac{(2o-25)^2}{25} + \frac{(3o-25)^2}{25} =$$

$$= 1 + 1 + 1 + 1 = 4$$

Si

$$\Upsilon = \sqrt{\frac{4}{1oo}} = o{,}2$$

Observaţie. Se vede din exemplele de mai sus - şi se poate uşor demonstra valabilitatea pe cazul general - că diferenţele ce apar la numărătorul fracţiilor în calculul lui χ^2 sînt egale în valoare absolută. Ele redau surplusul sau deficitul de indivizi în fiecare căsuţă a tabelului şi este clar că dacă într-o căsuţă avem un supliment de indivizi în celelalte două, de pe aceeaşi linie şi coloană, trebuie să avem acelaşi deficit.

Indicele lui Yule

Un alt indicator folosit pentru a ilustra şi măsura gradul de asociere într-un tabel de 2x2 a fost propus de statiscianul englez Yule şi este de forma următoare:

$$Q = \frac{k_{11}k_{22} - k_{12}k_{21}}{k_{11}k_{22} - k_{12}k_{21}}$$

Notațiile fiind cele din tabelul 1, se observă că la numărător apare diferența dintre produsul frecvențelor de pe diagonala principală și a celor de pe diagonala secundară, la numitor aceleași produse fiind însumate.

Proprietăți:

i) Indicele Q are intervalul de variație cuprins între -1 și +1.

ii) Din formulă rezultă imediat că dacă una din frecvențele k_{12} sau k_{21} se anulează, Q devine egal cu +1, iar dacă una dintre frecvențele k_{11} sau k_{22} este zero, Q va fi egal cu -1. In consecință, spre deosebirile de φ, Q atinge marginile intervalului de variație pentru implicația strictă univocă.

iii) Valoarea 0 este luată tot în cazul de independență a celor două caracteristici.

iv) Mărimea lui Q nu se schimbă dacă înmulțim sau împărțim o linie ori o coloană a tabelului cu un număr. Deci spre deosebire de φ, Q nu este sensibil față de structurile marginale. Această proprietate recomandă folosirea lui Q în cazul unor studii de eșantion cînd reprezentivitatea celor două clase ale unei caracteristici nu este asigurătă.

Exerciții:

1. Cazul tabelului 2.

$$Q = \frac{30x30 - 20x20}{30x30 + 20x20} = \frac{900 - 400}{900 + 400} = \frac{500}{1300} = 0,38$$

2. Cazul tabelului 4.

$$Q = \frac{30x60 - 0x10}{30x60 + 0x10} = \frac{30x60}{30x60} = 1$$

3. Cazul tabelului 7.

$$Q = \frac{9x2 - 88x26}{9x2 + 88x26} = \frac{18 - 2288}{18 + 2288} = \frac{-2270}{2306} = -0,98$$

4. Cazul tabelului 8.

$$Q = \frac{10x20 - 40x30}{10x20 + 40x30} = \frac{200 - 1200}{200 + 1200} = \frac{-1000}{1400} = -0,71$$

2. Cazul clasificărilor multiple

După ce am tratat destul de amănunțit situația particulară a clasificărilor dihotomice, vom trece în continuare să schițăm modalitatea de analiză a caracteristicilor ce divid populația în mai multe clase. Forma generală a tabelului de asociere - sau de contingență, cum mai poartă denumirea în astfel de cazuri - este cea redată în tabelul 10:

A \ B	B_1	B_2	B_3	...	B_j	...	B_t	Total
A_1	k_{11}	k_{12}	k_{13}	...	k_{1j}	...	k_{1t}	$k_{1.}$
A_2	k_{21}	k_{22}	k_{23}	...	k_{2j}	...	k_{2t}	$k_{2.}$
A_3	k_{31}	k_{32}	k_{33}	...	k_{3j}	...	k_{3t}	$k_{3.}$
.	.	.	.	.	.	.	.	.
A_i	k_{i1}	k_{i2}	k_{i3}	...	k_{ij}	...	k_{it}	$k_{i.}$
.	.	.	.	.	.	.	.	.
A_s	k_{s1}	k_{s2}	k_{s3}	...	k_{sj}	...	k_{st}	$k_{s.}$
Total	$k_{.1}$	$k_{.2}$	$k_{.3}$	...	$k_{.j}$	...	$k_{.t}$	n

Tabelul 10. Forma generală a tabelului de asociere (contingență)

Interpretarea notațiilor din tabelul 10 este următoarea:

A_1, A_2 etc. sînt simbolurile celor s clase ale lui A iar B_1, B_2 etc. sînt simbolurile celor t clase ale lui B; k_{ij}, unde i ia valori de la 1 la s iar j de la 1 la t, sînt frecvențele interioare ale tabelului, respectiv numărul indivizilor aflați simultan în clasa A_i și B_j ; în fine, $k_{1.}$, $k_{2.}$ $k_{3.}$, ...,$k_{s.}$ și $k_{.1}$, $k_{.2}$, $k_{.3}$,...., $k_{.t}$ sînt frecvențele marginale, adică frecvențele claselor lui A și B.

Vom spune că cele două caracteristici A şi B sînt <u>independente</u> dacă, oricare ar fi linia i şi coloana j, frecvenţele empirice k_{ij} coincid cu frecvenţele teoretice m_{ij} date de formula:

$$m_{ij} = \frac{k_{i.}\,k_{.j}}{n}$$

In caz contrar, cele două caracteristici se zice că sînt asociate.

Prin urmare, existenţa sau inexistenţa asocierii se poate depista prin compararea tuturor perechilor de valori k_{ij} şi m_{ij}, perechi care sînt în număr total de sxt. <u>Măsurarea gradului de asociere</u> se va face cu ajutorul unor indici care generalizează indicele φ de la cazul dihotomic, utilizînd mărimea χ^2. Din păcate, valoarea φ, dată de formula 5 nu mai posedă în cazul general proprietăţile unui indice, în sensul că ea poate depăşi unitatea, lucru care nu se întîmplă în cazul dihotomic. Va fi nevoie deci de o ajustare a valorii pentru a o aduce la o mărime subunitară. Există mai multe astfel de procedee, dintre care prezentăm două:

- Coeficientul lui Pearson

$$C = \pm\sqrt{\frac{\varphi^2}{1+\varphi^2}} = \pm\sqrt{\frac{\chi^2}{n+\chi^2}}$$

- Coeficientul lui Ciuprov

$$T = \pm\sqrt{\frac{\varphi^2}{\sqrt{(s-1)(t-1)}}} = \pm\sqrt{\frac{\chi^2}{n\sqrt{(s-1)(t-1)}}}$$

unde $\varphi^2 = \dfrac{\chi^2}{n}$ iar $\chi^2 = \sum_{i=1}^{s}\sum_{j=1}^{t}\dfrac{(k_{ij}-m_{ij})^2}{m_{ij}}$

Se observă că primul coeficient, C, nu poate atinge niciodată valoarea 1, deoarece numitorul este mai mare decît numărătorul, dar în cazul tabelelor de dimensiune mare el se apropie suficient de mult de unitate. Coeficientul lui Ciuprov ajunge la

valoarea 1, atunci cînd tabelul este pătratic - s = t - şi cînd toţi indivizii se plasează pe diagonala tabelului.

Atribuirea semnului plus sau minus coeficienţilor este o operaţie mai dificilă acum decît în cazul tabelelor de dimensiune 2x2. In mod riguros, semnul se poate atribui numai atunci cînd sînt îndeplinite simultan următoarele condiţii:

a) Tabelul este pătratic, deci s=t.
b) Ambele caracteristici sînt construite cu ajutorul unor scale ordinale.
c) Dacă clasele unei caracteristici sînt aşezate într-o anumită ordine, atunci rezultă logic o ordine şi pentru clasele celeilalte.

Dacă cele trei condiţii de mai sus sînt îndeplinite, atunci se acordă semnul plus în cazul în care indivizii se plasează pe diagonala principală şi semnul minus cînd ei se concentrează pe cea secundară. Dacă prima condiţie nu este îndeplinită, tabelul nu va avea căsuţe diagonale. In acest caz se poate totuşi ataşa un semn ca şi mai înainte, lărgind într-o anumită măsură sensul noţiunii de diagonală. Dacă nu sînt satisfăcute condiţiile b) şi c), atunci o eventuală ataşare a semnului trebuie însoţită de o explicaţie din care să rezulte sensul exact în care se poate vorbi de o asociere.

Dar pentru a înţelege mai bine cele spuse mai sus, precum şi felul cum se calculează indicatorii de asociere, să luăm cîteva exemple.

Exerciţiul 1. Să presupunem că într-o cercetare de opinie am cerut subiecţilor să aprecieze în termeni de favorabilitate două evenimente A şi B. Datele sînt cele din tabelul 11.

A \ B	Favorabil	Neutru	Nefavorabil	Total
Favorabil	20	30	150	200
Neutru	100	300	100	500
Nefavorab.	180	70	50	300
Total	300	400	300	1000

Tabelul 11.

Este limpede că în acest caz ne aflăm în situaţia cînd sînt îndeplinite toate cerinţele pentru acordarea semnului coeficinetului pe care urmează să-l calculăm. Tabelul este de dimensiune 3x3 şi dacă am ales pentru A prima clasă "favorabil" este normal s-o alegem şi pentru B tot pe aceea. Construim frecvenţele întîmplătoare m_{ij}:

60	80	60
150	200	150
90	120	90

şi calculăm:

$$\chi^2 = \frac{(20-60)^2}{60} + \frac{(30-80)^2}{80} + \frac{(150-60)^2}{60} + \frac{(100-150)^2}{150} + \frac{(300-200)^2}{200} + \frac{(100-150)^2}{150} + \frac{(180-90)^2}{90} + \frac{(70-120)^2}{120} + \frac{(50-90)^2}{90} =$$

$$= \frac{(-40)^2}{60} + \frac{(-50)^2}{80} + \frac{90^2}{60} + \frac{(-50)^2}{150} + \frac{100^2}{200} + \frac{(-50)^2}{150} + \frac{90^2}{90} + \frac{(-50)^2}{120} + \frac{(-40)^2}{90} =$$

$$= \frac{1600}{60} + \frac{2500}{80} + \frac{8100}{60} + \frac{2500}{150} + \frac{10000}{200} + \frac{2500}{150} + \frac{8100}{90} + \frac{2500}{120} + \frac{1600}{90} =$$

$$26{,}7 + 31{,}2 + 135 + 16{,}7 + 50 + 16{,}7 + 90 + 20{,}8 + 17{,}8 = 405$$

Prin urmare:

$$C = -\sqrt{\frac{405}{1000+405}} = -\sqrt{0{,}288} = -\,0{,}54$$

iar

$$T = -\sqrt{\frac{405}{1000\ \ (3-1)(3-1)}} = -\sqrt{\frac{405}{2000}} =$$

$$= -\ 0{,}2025 = -0{,}45$$

Semnul minus a fost acordat deoarece avem o concentrare pe diagonala secundară. Acest semn ne spune prin sine că aprecierile celor două evenimente au fost discordante, în general, în sensul că din-

trei cei care au apreciat unul ca favorabil mai mulţi au apreciat pe celălalt ca nefavorabil etc.

Exerciţiul 2. Să ne plasăm în aceeaşi situaţie ipotetică de mai sus numai să presupunem că pentru A am utilizat o scală cu 5 trepte (tab.12)

A \ B	Favorabil	Neutru	Nefavorabil	Total
Foarte favor.	0	1o	5o	6o
Favorabil	2o	2o	1oo	14o
Neutru	1oo	3oo	1oo	5oo
Nefavorabil	8o	5o	5o	18o
Foarte nef.	1oo	2o	0	12o
Total	3oo	4oo	3oo	1ooo

Tabelul 12.

Frecvenţele întîmplătoare vor fi:

18	24	18
42	56	42
15o	2oo	15o
54	72	54
36	48	36

Calculînd , se obţine

$$\chi^2 = 466{,}8$$

şi de aici

$$C = -0{,}56 \quad ; \quad T = -0{,}41$$

Semnul minus a putut fi acordat chiar în lipsa unei diagonale pro priu-zise, căci se vede că cei care apreciază B ca favorabil apreciază pe A mai degrabă în termeni negativi etc.

Exercițiul 3. Să presupunem că asociem o opinie (caracteristica B) cu clasa socială căreia aparține subiectul (A). Se obține tabelul 13.

A \ B	Favorabil	Neutru	Nefavorabil	Total
Muncitor	loo	2oo	3oo	6oo
Țăran	15o	loo	5o	3oo
Alte	5o	loo	5o	2oo
Total	3oo	4oo	4oo	11oo

Tabelul 13.

In cazul de față, caracteristica A este abordată cá o scală nominală și prin urmare nu are nici un sens să acordăm asocierii un semn. In astfel de cazuri nu este recomandabil să se treacă direct la calcularea coeficientului de asociere - cum am făcut în cazurile precedente - ci, trebuie analizat tabelul cu ajutorul frecvențelor (procentelor) la totalul unei caracteristici (cea care ar putea-o explica pe cealaltă). In situația noastră, vom calcula procente la totalurile lui A (tabelul 14).

A \ B	Favorabil	Neutru	Nefavor.	Total
Muncitor	16,7	33,3	5o,o	loo
Țăran	5o,o	33,3	16,7	loo
Alte	25,o	5o,o	25,o	loo
Total	27,3	36,4	36,4	loo

Tabelul 14. Frecvențele tabelului 13 exprimate procentual

Se vede acum foarte bine felul cum se structurează opiniile exprimate prin caracteristica B, în funcție de categoria socială a indivizilor studiați: muncitorii se concentrează spre "nefavorabil", țăranii spre "favorabil" iar restul spre "neutru".

Astfel devine limpede că un coeficient de asociere va reflecta aceste "concentrări" de opinii şi că este imposibil să-i acordăm un semn. Valorile coeficinetilor menţionaţi vor fi:

$$C = 0,36 \quad ; \quad T = 0,27$$

Observaţie. O analiză în termeni de procente, ca cea din exerciţiul 3, este recomandabil să fie făcută în toate cazurile deoarece ea ne ajută să vedem foarte clar sensul asocierii mai ales că în situaţiile reale nu se obţin cifre atît de rotund ca mai sus şi deci inspecţia primară a tabelului e mai dificilă.

3. Analiza multivariată

Sub titlul de analiză multivariată se înţelege, în domeniul metodologic al sociologiei, ansamblul de proceduri statisti menite să ducă la interpretarea relaţiilor complexe ce pot interni între mai multe caracteristici calitative. Metoda s-a impus aici mai ales în urma remarcabilei lucrări a lui E. Durkheim Le Suicide, a cărei influenţă s-a făcut simţită cu deosebită vigoare în rîndul sociologilor americani. Aceştia au încercat să descopere osatura metodologică a discursului durkheimian, reli fînd schemele generale ce pot fi folosite în analiza relaţiilor posibile dintre trei sau mai multe caracteristici, scheme care evidenţiază anumite cazuri tipice de asemenea relaţii. Mă voi opri pe scurt asupra celebrei - s-ar putea zice - scheme a lui P Lazarsfeld, aplicabilă în cazul a trei caracteristici dihotomice (1966).

Să presupunem că ne interesează influenţa unui factor A asupra altuia B. Primul pas îl constituie, evident, construirea unui tabel simplu de asociere, ca cele folosite în pragrafele precedente, din care se poate deduce existenţa sau lipsa unei asocieri între A şi B . Pasul următor va consta în introducerea unui al treilea factor C, numit, de regulă, variabilă test, despre care se poate presupune că va contribui la elucidarea relaţiei dintre A şi B. In esenţă, e vorba de a scinda populaţia noastră de n indivizi în două subpopulaţii, ce constituie efectivele celor două subclase ale lui C: C_1 şi C_2 , iar apoi, în cadrul fiecăreia

dintre aceste subpopulaţii, să se construiască tabele de asociere între A şi B . Dacă, de exemplu, în fiecare clasă a lui C, asocierea lui A cu B este aceeaşi ca şi asocierea pe ansamblul populaţiei, se poate trage concluzia că C nu manifestă nici o acţiune asupra relaţiei dintre cei doi factori şi, prin urmare, nu aduce nici un plus de informaţie în analiză. Dacă, dimpotrivă, există diferenţe între asocierile din cele două subpopulaţii - şi deci între fiecare subpopulaţie şi ansamblul - , atunci acţiunea lui C trebuie luată în considerare şi studiată. Or, tocmai aici şi-a adus Lazarsfeld contribuţia esenţială, evidenţiind unele "tipuri ideale" în mulţimea enormă de situaţii ce pot apare atunci cînd se asociază simultan trei factori.

Pentru o mai clară precizare a ideilor şi pentru a facilita înţelegerea procedurii lui Lazarsfeld, voi începe cu un exemplu concret, construit pe baza unor date fictive, tocmai pentru a ne plasa exact în una din situaţiile limită imaginate de sociologul american. Fie A factorul sex, cu clasele A_1 pentru bărbaţi şi A_2 pentru femei şi B o variabilă care să reprezinte participarea (B_1) sau neparticiparea (B_2) la activităţile sportive de masă. Să alegem ca variabilă test, C, faptul că persoanele în cauză efectuează (C_1) sau nu efectuează (C_2) munci casnice.

C / B / A	Efectuează munci casnice (C_1)			Nu efectuează munci casnice (C_2)			T O T A L		
	Participă (B_1)	Nu participă (B_2)	Tot	Participă (B_1)	Nu participă (B_2)	Tot.	Participă (B_1)	Nu participă (B_2)	Tot.
Bărbaţi (A_1)	1o	9o	1oo	2oo	2oo	4oo	21o	29o	5oo
Femei (A_2)	4o	36o	4oo	5o	5o	1oo	9o	41o	5oo
Total	5o	45o	5oo	25o	25o	5oo	3oo	7oo	1ooo

Tabelul 15. Distribuţia persoanelor după cele trei caracteristici, A, B şi C.

Tabelul 15 prezintă datele ipotetice obţinute. Se observă din ultima treime a tabelului că, pe ansamblul celor 1.000 de persoane investigate, există o asociere strînsă între sex şi participarea la acţiunile sportive. Bărbaţii iau parte în proporţie de 42% (210 din cei 500), pe cînd femeile numai de 18% (90 din cele 500). Privind acum primele două părţi ale tabelului, în care sînt separate cele două loturi după C - 500 persoane care efectuează munci casnice şi 500 care nu efectuează -, vedem imediat că diferenţa pe sexe dispare: în primul caz, atît participarea bărbaţilor cît şi a femeilor este de 10% (10 din 100, respectiv 40 din 400), iar în al doilea de 50% (200 din 400 şi, respectiv, 50 din 100).

Din datele prezentate (cu repetarea precizării că ele sînt absolut fictive), rezultă că numai antrenarea în munci casnice, diferită pentru cele două sexe, este responsabilă de inegala participare a bărbaţilor şi femeilor la acţiunile sportive. Dacă am fi analizat doar asocierile simple între cei trei factori (A cu B, A cu C, B cu C), s-ar fi observat că bărbaţii sînt mai puţin prinşi în munci casnice, apoi că cei ce efectuează asemenea munci au o participare mai redusă la activitatea sportivă, ceea ce implică imediat asocierea existentă între sex şi practicarea sportului. Dar această implicaţie este una slabă şi ea nu este capabilă să ne indice intensitatea asocierii lui A cu B şi, cu atît mai puţin, să ne conducă la concluzia că întreaga asociere a lui A cu B se datorează variabilei test C. Numai analiza tabelului de asociere simultană a celor trei factori poate pune cu claritate în evidenţă această concluzie.

Dacă notăm cu $\tau(A,B)$, $\tau(A,C)$, $\tau(B,C)$ coeficienţii de asociere între cele trei perechi de caracteristici şi cu $\tau(A,B;C_1)$ şi $\tau(A,B;C_2)$ asocierea dintre A şi B în cele două clase ale lui C, deci coeficienţii parţiali de asociere ai lui A cu B cînd C este menţinut constant, atunci se poate scrie o egalitate fundamentală, datorată lui Lazarsfeld:

$$\tau(A,B) = \alpha\cdot\tau(A,B;C_1) + \beta\cdot\tau(A,B;C_2) + \gamma\cdot\tau(A,C)\cdot\tau(B,C)$$

unde α, β şi γ sînt trei coeficienţi pozitivi.

In exemplul prezentat mai sus, se vede că vom avea:

$$\mathcal{C}(A,B;C_1) = \mathcal{C}(A,B;C_2) = 0$$

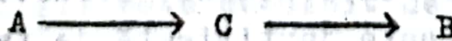
deci relația lui Lazarsfeld va avea forma:

$$\mathcal{C}(A,B) = 0 + 0 + \gamma \cdot \mathcal{C}(A,C) \cdot \mathcal{C}(B,C)$$

care caracterizează tipul numit de sociologul american "elaborare prin relații marginale (M)" iar de către R.Boudon (1967) "structură cu efect de regresie". Acest tip se descompune, la rîndul său, în alte două subtipuri, după cum variabila test este anterioară lui A, cazul MA sau intermediară lui A și B, cazul MI.

Distincția dintre situațiile MA și MI nu este deloc convențională, ci are implicații serioase în puterea explicativă a modelului celor trei variabile analizate. Intr-adevăr, se vede că, în general, într-o situație de tipul M, relația statistică a lui A cu B se datorează în exclusivitate factorului C. Dacă C se plasează între A și B, cum se întîmpla în exemplul nostru, atunci putem spune că A nu are o influență directă asupra lui B și că influența sa se transmite prin intermediul lui C. Influența sexului asupra participării la activitatea sportivă se materializează în diferența de participare la activitățile casnice. Analog este și exemplul pe care-l dă Lazarsfeld: diferența dintre tineri și vîrstnici în audiția la radio a emisiunilor religioase - în favoarea celor din urmă - se explică prin intervenția factorului instrucție, inegal distribuit în cele două grupe de vîrstă, intervenție care se concretizează prin faptul că la fiecare nivel de instrucție asocierea dintre vîrstă și ascultarea emisiunilor religioase dispare.

Schematic, situația MI se poate reprezenta astfel:



$$A \longrightarrow C \longrightarrow B$$

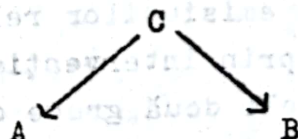
In cazul în care variabila test este anterioară factorului explicativ A, asocierea statistică dintre A și B este absolut străină unei relații de influențare, atunci cînd sînt satisfăcute condițiile tipului M (anularea asocierilor parțiale). Ea se datorează și acum factorului test, C, numai că el fiind anterior și lui A și lui B, nu mai e vorba de o legătură propriu-zisă, nici

măcar mediată, între A și B. Dacă transpunem în limbaj cauzal situația MA, atunci putem spune că relația lui A cu B rezultă din faptul că ambele fenomene sînt efectele lui C. Exemple pentru această situație se pot imagina destule, dar prefer să reproduc două care au devenit deja clasice în manualele de statistică:

a) Se constată, pentru o mulțime de localități, o asociere strînsă pozitivă între numărul de berze din localitate (factorul A) și rata de natalitate în localitatea respectivă (B). Legătura dintre A și B este absolut iluzorie, fapt care se poate demonstra dacă se introduce ca variabilă test (C) tipul de localitate (oraș-sat). In mediul rural, berzele sînt mai numeroase decît la orașe iar natalitatea este de asemenea, mai mare.

b) Se constată, la fel, o asociere, foarte strînsă între cantitatea de apă aruncată de pompieri în cursul unui incendiu (factorul A) și gravitatea pagubelor produse de incendiu (B). Evident că nici această asociere nu reflectă o relație de determinare de la A spre B; un al treilea factor, anterior lui A și B, pe care-l putem denumi "gravitatea incendiului" explică atît pagubele produse cît și volumul apei utilizată de pompieri.

Pentru situația MA, schema grafică este următoarea:



Al doilea tip principal prezentat de Lazarsfeld este cel în care asocierea dintre A, variabila independentă, și C variabila test, este nulă. El numește acest caz "elaborare prin relații parțiale (P)" iar Boudon preferă expresia de "structură cu efect de interacțiune" Prima denumire este justificată de faptul că, datorită anulării lui $\mathcal{C}(A,C)$, ecuația lui Lazarsfeld devine:

$$\mathcal{C}(A,B) = \alpha \cdot \mathcal{C}(A,B;C_1) + \beta \cdot \mathcal{C}(A,B;C_2)$$

adică asocierea lui A cu B apare doar ca o rezultantă a asocierilor parțiale: $\mathcal{C}(A,B;C_1)$ și $\mathcal{C}(A,B;C_2)$. A doua subliniază faptul că tocmai aceste asocieri parțiale sînt indicatorii de măsurare a efectului interacțiunii lui A și C în determinarea lui B.

Să luăm şi pentru cazul de faţă un exemplu, de asemenea cu date fictive, reprezentat în tabelul 16.

C	Instrucţie superioară (C_1)			Instrucţie inferioară (C_2)			TOTAL		
A \ B	Interes (B_1)	Dezinteres (B_2)	Tot.	Interes (B_1)	Dezinteres (B_2)	Tot.	Interes (B_1)	Dezinteres (B_2)	Tot.
Bărbaţi(A_1)	13o	2o	15o	25o	1oo	35o	38o	12o	5oo
Femei (A_2)	12o	3o	15o	1oo	25o	35o	22o	28o	5oo
Total	25o	5o	3oo	35o	35o	7oo	6oo	4oo	1ooo

Tabelul 16. Un exemplu de elaborare prin relaţii parţiale

Fie A variabila sex, B - interesul pentru probleme de politică externă şi variabila test, C - nivelul de instrucţie şcolară. Cifrele din tabelul 16 arată că sexul nu este asociat cu instrucţia: 15o de bărbaţi din cei 5oo şi 15o de femei tot din 5oo au nivelul superior de instrucţie. Dimpotrivă, atît sexul cît şi instrucţia se află într-o relaţie statistică destul de strînsă cu interesul pentru politica externă: 38o de bărbaţi se interesează, faţă de numai 22o de femei, din acelaşi efectiv total de 5oo; 5/6 (25o din 3oo) dintre cei cu instrucţie superioară, faţă de numai jumătate (35o din 7oo) dintre cei cu nivel şcolar scăzut. In atari condiţii este deosebit de importantă informaţia adusă de asocierile parţiale. Este imediat vizibil că în subeşantionul celor cu instrucţie superioară (clasa C_1) asocierea dintre sex şi interes este deosebit de modestă, aproape nulă: bărbaţii se interesează în procent de 86,7% (13o din 15o) iar femeile în proporţie de 80% (12o din 15o). Din contră, în celălalt subgrup (C_2) proporţia bărbaţilor interesaţi de politică depăşeşte spectaculos pe cea a femeilor: 71,4 % (25o din 35o) faţă de numai 28,6% (1oo din 35o). Prin urmare, dacă datele prezentate ar fi reale,

am putea afirma că influenţa sexului asupra interesului pentru probleme politice se materializează doar în interacţiune cu educaţia şcolară, în sensul că diferenţele dintre sexe se şterg la nivelurile superioare de instrucţie, rămîn clare doar la cele inferioare.

Exemplul acesta poate fi interpretat ca reflectînd o situaţie experimentală cu grup de control. Fenomenul A apare drept stimul, C drept dispoziţie iar B răspuns. Intr-un astfel de limbaj se înţelege mai uşor de ce diferenţa de asociere dintre A şi B în cele două clase ale lui C poate fi numită efect al interacţiunii lui A cu C. (Stimului interacţionează cu dispoziţia în producerea răspunsului).

Există şi de data aceasta două subcazuri. Unul, ca cel reflectat de exemplul de mai sus. cînd C este intermediar lui A şi B, este notat de Lazarsfeld cu simbolul PI; celălalt, notat cu PA, corespunde situaţiei în care C este anterior lui A. Distincţia dintre PI şi PA nu este atît de importantă ca cea dintre MI şi MA, în cele două cazuri P, schimbîndu-se doar rolurile variabilei test şi a celei independente.

Interpretarea în limbajul experimentului se datorează aici faptului că asocierea dintre A şi C este nulă. Intr-adevăr dacă cei care posedă calitatea A_1 constituie grupul experimental iar cei care n-o posedă (cei din A_2), grupul de control, lipsa de asociere dintre A şi C nu înseamnă altceva decît că, din punctul de vedere al comportamentului faţă de C, grupurile experimental şi de control sînt identice. Or, aceasta era una dintre supoziţiile fundamentale ale experimentului cu grup de control. Dacă situaţia reală nu ne conduce la un astfel de raport între A şi C, ea poate fi "forţată" să se încadreze într-o asemenea schemă, alegînd convenabil indivizii din A_2 sau din A_1. In cazul constru, de pildă, dacă bărbaţii şi femeile n-ar fi avut aceeaşi distribuţie după şcolaritate, s-ar fi putu alege un lot de femei (sau de bărbaţi) care să aibă o structură şcolară identică cu a bărbaţilor (sau femeilor). Desigur că atunci corelaţiile ansamblului eşantionului φ(A,B), φ(A,C) şi φ(B,C) ar fi fost denaturate într-o oarecare măsură, dar, la urma urmei, ele nu prezintă interes ca valori exacte. Diferenţa de comportament a indivizilor din cele două clase ale lui A în raport cu variabila B s-ar menţine (putînd fi măsu-

rată prin δ_A) iar interacţiunea dintre A şi C în determinarea lui B ar ieşi mai clar în evidenţă prin anularea lui $\Upsilon(A,C)$

In general, se cuvine observat că situaţiile reale rareori se suprapun peste vreunul din tipurile ideale descrise, sau pot fi aduse la acestea. De cele mai multe ori, sîntem conduşi să analizăm tabele fără a putea măcar spunecă ele se apropie de tipurile lui Lazarsfeld, ceea ce nu diminuează cu nimic importanţa distincţiilor teoretice de mai sus. Intr-adevăr, din analiza precedentă rezultă unele învăţăminte de care trebuie să ţinem seama atunci cînd analizăm tabele empirice efectiv rezultate dintr-o cercetare:

a) Dacă vom găsi că mărimile asocierilor parţiale $\Upsilon(A,B;C_1)$ şi $\Upsilon(A,B;C_2)$ sînt aproximativ egale şi au acelaşi semn cu asocierea de ansamblu $\Upsilon(A,B)$,atunci înseamnă că o parte din asocierea lui A şi B trebuie pusă pe seama lui C, iar această parte este cu atît mai mare cu cît asocierea totală diferă mai mult de asocierile parţiale. La limită, putem ajunge în două situaţii: cînd asocierile parţiale sînt nule (cazul II) - şi deci întreaga corelaţie (A,B) se poate pune pe seama lui C - sau cînd asocierile parţiale coincid cu asocierea totală perfect:

$$\Upsilon(A,B;C_1) = \Upsilon(A,B;C_2) = \Upsilon(A,B)$$

şi atunci introducerea factorului C în analiză este superfluă, deoarece acesta nu afectează cu nimic raportul lui A cu B. In toate situaţiile, poziţia lui C (înaintea lui A sau între A şi B) este foarte importantă pentru interpretarea relaţiilor dintre variabile.

b) In cazul în care asocierile parţiale diferă semnificativ una de cealaltă, atunci, indiferent de faptul că $\Upsilon(A,C)$ este sau nu egal cu zero, înseamnă că că ne găsim în situaţia de interacţiune. Cu cît diferenţa este mai mare cu atît interacţiunile sînt mai puternice şi nu este exclus să întîlnim situaţiile reale în care semnele celor două asocieri parţiale să fie contrare. Ne-am fi găsit într-o asemenea situaţie dacă, de pildă, în exemplul precedent, la nivelul de instrucţie superior, femeile ar fi fost cele care se interesează mai mult de politică, iar la cel inferior, bărbaţii.

Este evident că, dacă cele două asocieri parţiale sînt de semne contrare, se poate ajunge - prin "raportul de forţe" al ponderilor diferitelor clase ale variabilelor - la situaţia în care asocierea totală (A,B) să fie nulă, în condiţii deci de influenţă a lui A asupra lui B. Prin urmare, nici existenţa unei asocieri între A şi B nu reprezintă o garanţie a unei legături între factorii respectivi (vezi cazul MA) şi nici lipsa unei asocieri nu este o dovadă a independenţei reale a celor doi factori.

Am pomenit mai înainte despre apropierea dintre tipul P şi o situaţie experimentală; nu este posibil să nu se remarce faptul că de fapt orice analiză multivariată încearcă o rafinare a datelor de observaţie în scopul apropierii de condiţiile experimentului. Este evident că introducerea variabilei test şi analiza relaţiilor dintre factorul "independent" şi cel "dependent", în clasele variabilei test, nu înseamnă altceva decît controlul lui C. Dar, după cum am spus cînd am vorbit despre experiment, controlul trebuie să se aplice tuturor factorilor care pot influenţa raportul celor două variabile de bază. Această idee ne conduce la extinderea procedurii analizei multivariate, extindere ce se poate realiza pe două căi: fie prin introducerea succesivă a mai multor variabile test, fie prin introducerea simultană a acestora. Desigur, că din punctul de vedere al rezultatelor de cunoaştere, este preferabilă această din urmă variantă, deoarece ea realizează mai fidel una din condiţiile experimentale (controlul simultan al factorilor), dar, din păcate, practic o astfel de analiză este greu de realizat. In adevăr, chiar şi pentru cazul cel mai simplu al caracteristicilor dihotomice, numărul căsuţelor interioare - deci fără a mai socoti toate totalizările parţiale ce se pot face - creşte foarte repede: pentru 2 variabile test - 16 căsuţe, pentru 3 - 32 căsuţe, pentru 4 - 64 etc. Cînd clasificarea nu este dihotomică, cifrele de mai sus devin ameţitor de mari. Presupunînd, de pildă, că fiecare variabilă are cinci clase, un tabel cu o variabilă test va avea 125 de căsuţe interioare, unul cu 2 variabile test - 625 de căsuţe, cu 3 variabile test - 3125 ş.a.m.d. Creşterea numărului de căsuţe în tabel îngreunează analiza ducînd-o pînă la urmă la imposibilitate, din două motive. Mai întîi, este clar că nici o minte, oricît de versată ar fi în analize

statistice, nu poate "vedea" clar ce se petrece într-un tabel de asemenea dimensiuni. Apoi, şi ceea ce este mai important , aceste tabele "mamut" trebuie să conţină cifre în fiecare căsuţă iar frecvenţele respective trebuie să fie suficient de mari pentru a putea face comparaţii valide între diferitele rubrici ale tabelului. De aici rezultă necesitatea unor eşantioane enorme, de care sociologul rareori dispune.

Voi prezenta, în încheiere, un exemplu de analiză multivariată cu cifre reale, exemplu ce mai are şi particularitatea că numărul claselor este variabil în cele trei caracteristici. La un examen de admitere în învăţămîntul superior - vezi T.Rotariu (1974) - s-a urmărit, printre altele, construcţia caracteristicilor de mai jos:

- <u>Originea socială a candidaţilor</u> (<u>A</u>), distingîndu-se trei categorii:

  - A_1 - intelectuali
  - A_2 - categorii intermediare
  - A_3 - muncitori şi ţărani

- <u>Pregătirea şcolară din liceu</u> (C), formîndu-se patru clase, în funcţie de notele de la bacalaureat şi din ultima clasă de liceu:

  - C_1 - medii între 9 şi lo
  - C_2 - " " 8 şi 8,99
  - C_3 - " " 7 şi 7,99
  - C_4 - " sub 7

- <u>Reuşita la admitere</u> (<u>B</u>) variabila dependentă, cu două clase:

  - B_1 - admis
  - B_2 - respins

Pregătire școlară(C)	C_1			C_2			C_3			C_4		
Originea (A)	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
% de reușiți (B)	7o,4	63,0	57,4	54,5	42,6	37,8	31,1	25,9	21,4	2o,7	16,0	15,8

Tabelul 17. Influența pregătirii școlare și originii sociale asupra reușitei candidaților la admiterea în învățămîntul superior.

Tabelul 17 concentrează întreaga informație necesară analizei, cu excepția corelațiilor globale, care sînt toate pozitive. Plecînd de la asocierea dintre A și B, în sensul că fiii de intelectuali reușesc în procent mai ridicat, s-a pus problema dacă factorul intermediar C transmite sau nu toată influența lui A asupra lui B. Datele noastre ne arată că nu ne găsim în cazul de elaborare prin relații marginale, MI, deoarece la fiecare nivel de pregătire școlară se menține o asociere între origine și reușită, prin urmare există o cale de influențare ce nu trece prin C.

A doua constatare ce se impune din analiza datelor tabelului este existența unei interacțiuni între originea socială și pregătirea școlară în influențarea reușitei la admitere: pe măsură ce coborîm pe scara notelor - de la C_1 la C_4 -, efectele directe ale originii sociale se diminuează.

Cap. X. CORELATIA

Problematica ce constituie conţinutul acestui capitol este similară celei expuse în cazul asocierii, cu deosebirea că, de această dată, ne vom referi la caracteristici cantitative. Prin urmare, noţiunea de corelaţie înlocuieşte, pentru variabilele statistice, pe cea de asociere, folosită spre a ilustra interdependenţa caracteristicilor calitative.

S-ar părea că vorbind aici despre corelaţie ar trebui să repetăm în mare cele spuse cu ocazia prezentării asocierii, adăugînd ceva în plus din cauza proprietăţilor specifice variabilelor (clasele acestora sînt caracterizate de valori numerice). S-ar părea deci că, în esenţă, corelaţia nu este altceva decît asocierea aplicată la variabile. Lucrurile nu stau însă deloc aşa; este drept că între cele două noţiuni există o înrudire foarte strînsă dar ele nu trebuie totuşi confundate sau, mai corect spus, corelaţia nu este o speţă (un caz particular) de asociere. Este evident că, din faptul că variabilele pot fi privite drept caracteristici calitative, făcînd abstracţie de valori, rezultă imediat că tot ce s-a vorbit în capitolul precedent despre acestea din urmă este valabil şi aplicabil variabilelor. Adică noţiunea de asociere - respectiv independenţă - se poate folosi şi la variabile fără nici o restricţie.

Calea prin care am ajuns la definirea asocierii nu se petrează însă la dezvoltări fructuoase pentru variabile - datorită faptului că nu permite introducerea în calcule a valorilor - şi de aceea o vom părăsi, alegînd o alta care ne va conduce la definirea corelaţiei. Se va reveni apoi spre a compara cele două noţiuni şi a ilustra succint asemănările şi deosebirile dintre ele.

1. Convarianţă, coeficient de corelaţie

Fie X şi Y două variabile ce iau valorile $x_1, x_2, \dots, x_s$ şi, respectiv, $y_1, y_2, \dots, y_t$, valori scrise în ordinea lor crescătoare. Să notăm cu $k_{1.}, k_{2.}, \dots, k_{s.}$ şi cu $k_{.1}, k_{.2}, \dots, k_{.t}$ frecvenţele cu care sînt luate aceste valori iar cu k_{ij} $(i=1,2,\dots,s;\ j=1,2,\dots,t)$ numărul indivizilor care iau simultan valorile x_i şi y_j.

Forma generală a tabelului de corelaţie (tabelul 1) este similară celei a tabelului de asociere (tabelul 1o, cap.9)

Să luăm un individ oarecare din populaţia noastră şi să presupunem că, referitor la variabila X, el primeşte valoarea x_i iar, referitor la Y, pe y_j. Dacă $\bar{x}$ şi $\bar{y}$ sînt valorile medii ale celor două variabile, putem construi diferenţele:

$$x_i - \bar{x} \text{ şi } y_j - \bar{y}$$

X \ Y	y_1	y_2 ...	y_j ...	y_t	Total
x_1	k_{11}	k_{12} ...	k_{1j} ...	k_{1t}	$k_{1.}$
x_2	k_{21}	k_{22} ...	k_{2j} ...	k_{2t}	$k_{2.}$
...	...	...	...	...	...
x_i	k_{i1}	k_{i2} ...	k_{ij} ...	k_{it}	$k_{i.}$
...	...	...	...	...	...
x_s	k_{s1}	k_{s2} ...	k_{sj} ...	k_{st}	$k_{s.}$
Total	$k_{.1}$	$k_{.2}$ ...	$k_{.j}$ ...	$k_{.t}$	n

Tabelul 1. Forma generală a tabelului de corelaţie

Dacă una dintre diferenţe este pozitivă, atunci individul în cauză se află plasat, prin valoarea luată, în dreapta mediei şi, dacă e negativă, în stînga. De exemplu:

$$x_1 \quad x_2 \quad \ldots \quad x_i \quad \ldots \quad \bar{x} \quad \ldots \quad x_s$$

diferenţă negativă

In raport cu ambele variabile, pentru un acelaşi individ putem avea patru situaţii:

a) $\begin{cases} x_i - x \geqslant 0 \\ y_i - y \geqslant 0 \end{cases}$ b) $\begin{cases} x_i - x \leqslant 0 \\ y_j - y \leqslant 0 \end{cases}$

c) $\begin{cases} x_i - x \geqslant 0 \\ y_j - y \leqslant 0 \end{cases}$ d) $\begin{cases} x_i - x \leqslant 0 \\ y_j - y \geqslant 0 \end{cases}$

Primele două ne arată o concordanţă în plasarea indivizilor pe scala valorilor lui X şi Y, în raport cu mediile acestora, pe cînd ultimele o discordanţă. Cele patru cazuri se pot, în această idee, sintetiza în două. Intr-adevăr, prin efectuarea produselor diferenţelor, în cazurile a) şi b) se va obţine o mărime pozitivă (eventual nulă), iar pentru c) şi d) una negativă (sau nulă):

$$\alpha)\quad (x_i - \bar{x})(y_i - y) \geqslant 0 \quad \text{- concordanţă}$$
$$\beta)\quad (x_i - \bar{x})(y_j - y) \leqslant 0 \quad \text{- discordanţă}$$

Să ne imaginăm, în continuare, că vom efectua aceste produse pentru fiecare individ din populaţie, încercînd să vedem care situaţie este predominantă: cea de concordanţă sau cea de discordanţă. Desigur că nu este suficient să numărăm cazurile de un tip şi de celălalt şi să facem bilanţul, dat fiind că aceste produse nu au numai un <u>semn</u> dar şi o <u>valoare absolută</u>, care trebuie luată în considerare. Se pune deci problema de a găsi o valoare unică destinată să caracterizeze, într-o manieră sintetică, felul cum se "comportă" ansamblul indivizilor simultan în raport cu cele două variabile. Adică să arate în ce <u>măsură</u> (şi termenul de măsură este folosit aici în sensul său propriu) indivizii care iau valori mari (mici) pentru X iau valori mari (mici) şi pentru Y. Or, un indicator potrivit pentru o atare măsură este <u>media aritmetică a produselor</u> efectuate pentru toţi indivizii.

Privind tabelul de corelaţie, vedem că nu este nevoie să facem <u>n</u> asemenea produse. De pildă,

$$(x_1 - \bar{x})(y_1 - \bar{y})$$

este comun mai multor indivizi, şi anume pentru toţi cei aflaţi în căsuţa din colţul stîng şi de sus al tabelului, în număr de k_{11}.

Suma acestui produs pentru toţi cei k_{11} indivizi va fi evident:

$$k_{11}(x_1 - \bar{x})(y_1 - \bar{y})$$

In general, vom ajunge, pentru calcularea mediei, la o sumă de termeni de forma:

$$k_{ij}(x_i - \bar{x})(y_j - \bar{y})$$

unde indicele <u>i</u> ia valori de la 1 la <u>s</u>, iar indicele <u>j</u> de la 1 la <u>t</u>, deci la o sumă de s.t termeni (numărul liniilor înmulţit cu numărul coloanelor sau numărul căsuţelor interioare ale tabelului), sumă care finalemente trebuie împărţită la <u>n</u>, numărul total de indivizi. Mărimea astfel obţinută se numeşte <u>covarianţa</u> variabilelor X şi Y şi se notează cu Cov(X,Y). Prescurtat, ea se scrie:

$$\mathrm{Cov}(X,Y) = \frac{\sum_{i=1}^{s}\sum_{j=1}^{t} k_{ij}(x_i - \bar{x})(y_j - \bar{y})}{n}$$

In cazul în care Cov (X,Y) este nulă, vom spune că variabilele X şi Y sînt <u>necorelate</u>, iar dacă este diferită de zero, că ele sînt corelate. Cînd Cov(X,Y) < 0, vom avea o corelaţie negativă (însemnînd că discordanţele de comportament prevalează asupra concordanţelor) şi cînd Cov(X,Y) > 0, corelaţia se zice pozitivă.

Prin urmare, <u>noţiunea de corelaţie</u> a variabilelor statistice se poate introduce cu ajutorul <u>mărimii covarianţă</u>, a cărei semnificaţie am sugerat-o în rîndurile de mai sus. Deşi covarianţa indică existenţa sau inexistenţa corelaţiei, precum şi sensul acesteia (cînd există), ea nu poate juca totuşi rolul unui <u>coeficient</u> de corelaţie pentru motivul că depinde de unităţile de măsură ale celor două variabile şi nu are limite fixe de variaţie, deci nu îndeplineşte două dintre condiţiile fundamentale ale oricărui coeficient.

Există o procedură foarte simplă de a înlătura neajunsurile menţionate. Este suficient să se împartă covarianţa la abaterile standard ale lui X şi Y şi se va obţine coeficientul dorit. Prin urmare, formula <u>coeficientului de corelaţie</u>, notat cu <u>r</u>, este:

$$r = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

sau

$$r = \frac{\sum_{i=1}^{s}\sum_{j=1}^{t} k_{ij}(x_i-\bar{x})(y_j-\bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \qquad (1)$$

Se vede că prin împărţirea cu $\sigma_x \cdot \sigma_y$ dispare dimensiunea mărimii şi se poate demonstra că:

$$-1 \leq r \leq +1$$

Pentru calcularea coeficientului de corelaţie cu ajutorul formulei (1), trebuie efectuate succesiv următoarele operaţii:

1. Se calculează valorile medii, $\bar{x}$ şi $\bar{y}$, ale celor două variabile.

2. Se calculează abaterile standard σ_x şi σ_y.

3. Se scade $\bar{x}$ din fiecare valoare x_i şi $\bar{y}$ din fiecare y_j.

4. Se efectuează produsele $k_{ij}(x_i - \bar{x})(y_j - \bar{y})$, adică, urmărind frecvenţele interioare ale tabelului, k_{ij}, acestea se înmulţesc cu diferenţele obţinute la punctul 3, diferenţe ce caracterizează fiecare linie şi coloană a tabelului.

5. Se efectuează suma algebrică (cu semn) a mărimilor de la punctul 4, obţinîndu-se astfel numărătorul coeficientului de corelaţie.

6. Se calculează numitorul: $n \sigma_x \sigma_y$

7. Se efectuează împărţirea .

Se observă că cele mai mari dificultăţi apar la punctele 4 şi 5, deoarece, de regulă, valorile medii calculate sînt sub formă de numere cu zecimale şi, prin urmare, şi diferenţele vor fi la fel, ceea ce conduce la un volum mare de muncă pentru efectuarea înmulţirirlor. De exemplu, dacă valorile lui X sînt 1,2,3,4 şi 5 iar media iese 2,75 , atunci diferenţele vor fi:

-1,75; -0,75; +0,25; +1,25 şi 2,25. Pentru Y se vor întîmpla foarte probabil lucruri similare şi atunci înmulţirea acestor numere cu zecimale nu este o treabă foarte plăcută chiar dacă folosim un minicalculator de birou.

Pentru uşurarea calculelor se pot folosi şi alte formule, echivalente cu (1). Dacă efectuăm operaţiile de la numărător în formula (1) - deschidem parantezele - vom ajunge la următoarea formă a coeficientului de corelaţie, reducînd termenii asemenea:

$$r = \frac{\sum_{i=1}^{s} \sum_{j=1}^{t} k_{ij} x_i y_j - n\bar{x}\bar{y}}{n \cdot \sigma_x \cdot \sigma_y} \qquad (2)$$

Folosirea formulei (2) - în care se vede că frecvenţele k_{ij} trebuie înmulţite direct cu valorile corespunzătoare x_i şi y_j iar mediile apar doar într-o singură operaţie de înmulţire - se recomandă atunci cînd valorile x_i şi y_j sînt mici, deci cînd produsele $k_{ij}x_iy_j$ nu ne duc la cifre astronomice.

O altă modalitate de simplificare a calculelor este cea care face apel la utilizarea unor valori auxiliare, a şi b în locul valorilor medii $\bar{x}$ şi $\bar{y}$. Procedura este identică cu cea expusă cu ocazia prezentării abaterii standard şi, de fapt, aceste valori vor fi folosite şi aici în calculul celor două abateri standard dar şi la suma dublă de la numărător. Formula poate fi scrisă astfel:

$$r = \frac{\sum_{i=1}^{s} \sum_{j=1}^{t} k_{ij}(x_i - a)(y_j - b) - n\, d_x d_y}{n \cdot \sigma_x \sigma_y} \qquad (3)$$

unde

$$d_x = \bar{x} - y$$
$$d_y = \bar{y} - b$$

Valorile a şi b se aleg şi acum de aşa manieră încît să se obţină cifre cît mai mici pentru diferenţele $x_i - a$ şi $y_j - b$. Dacă a este chiar una dintre valorile lui X şi b una dintre cele ale lui Y, atunci, în plus, pentru o linie şi o coloană produsele vor fi nule.

Voi ilustra cu un exemplu mersul calculelor atunci cînd se utilizează formula (3). Să presupunem că într-o cercetare pe un lot de 1000 de muncitori am introdus două variabile: "vîrsta" (X) şi "retribuţia" (Y). Ambele variabile sînt cu intervale de grupare, fiecare interval fiind apoi caracterizat de o valoare numerică. Distribuţia indivizilor după cele două variabile este dată în tabelul 2, în care sînt trecute şi valorile caracteristice intervalelor de grupare.

X \ Y	2500 lei	2750 lei	3000 lei	3250 lei	3500 lei	Total
20 ani	50	30	15	5	-	100
30 ani	25	70	30	15	10	150
40 ani	15	50	200	60	25	350
50 ani	10	30	40	100	20	200
60 ani	-	20	15	70	95	200
Total	100	200	300	250	150	1000

Tabelul 2. Corelaţia între vîrsta (X) şi retribuţia (Y) ale unui lot de 1000 muncitori (date fictive)

Se observă că, pe măsură ce avansăm pe scara vîrstei, ponderea indivizilor cu trepte mai înalte de retribuţie creşte. Prin urmare, tabelul 2 redă un caz tipic de corelaţie pozitivă iar calculul coeficientului va arăta intensitatea acestei corelaţii.

Alegem următoarele valori ajutătoare:

a = 40 ani
b = 3000 lei

şi aranjăm datele pentru calcularea mediilor şi abaterilor standard sub forma cunoscută:

Pentru X:

x_i	$x_i - a$	$k_{i.}$	$k_{i.}(x_i-a)$	$(x_i-a)^2$	$k_{i.}(x_i - a)^2$
2o	-2o	1oo	-2.ooo	4oo	4o.ooo
3o	-1o	15o	-1.5oo	1oo	15.ooo
4o	0	35o	-	-	-
5o	1o	2oo	2.ooo	1oo	2o.ooo
6o	2o	2oo	4.ooo	4oo	8o.ooo
Total		1ooo	2.5oo		155.ooo

Rezultă imediat:

$d_x = 2,5$ $\quad d_x^2 = 6,25$ $\quad \bar{x} = 42,5$

$s_x^2 = 155$ $\quad \sigma_x^2 = 148,75$ $\quad \sigma_x = 12,2$

Pentru Y:

y_j	$y_j - b$	$k_{.j}$	$k_{.j}(y_j - b)$	$(y_j - b)^2$	$k_{.j}(y_j - b)^2$
2.5oo	-5oo	1oo	-5o.ooo	25o.ooo	25.ooo.ooo
2.75o	-25o	2oo	-5o.ooo	62.5oo	12.5oo.ooo
3.ooo	0	3oo	-	-	-
3.25o	25o	25o	62.5oo	62.5oo	15.625.ooo
3.5oo	5oo	15o	75.ooo	25o.ooo	37.5oo.ooo
Total		1ooo	37.5oo		9o.625.ooo

Si deci:

$d_y = 37,5$ $\quad d_y^2 = 1406,25$ $\quad \bar{y} = 3.037,5$

$s_y^2 = 9o.625$ $\quad \sigma_y^2 = 89.218,75$ $\quad \sigma_y = 298,7$

Trecem apoi la calcularea sumei de la numărător. In schema de calcul ce urmează am aşezat pe prima coloană frecvenţele interioare ale tabelului de corelaţie, urmărindu-le pe linii.

Pentru produsele finale am construit două coloane, pentru a trece separat valorile pozitive şi negative.

k_{ij}	$x_i - a$	$y_j - b$	$k_{ij}(x_i - a)(y_j - b)$ +	$k_{ij}(x_i - a)(y_j - b)$ -
50	-20	-500	500.000	
30	-20	-250	150.000	
15	-20	0		
5	-20	250		25.000
0	-20	500		
25	-10	-500	125.000	
70	-10	-250	175.000	
30	-10	0		
15	-10	250		37.500
10	-10	500		50.000
15	0	-500		
50	0	-250		
200	0	0		
60	0	250		
25	0	500		
10	10	-500		50.000
30	10	-250		75.000
440	10	0		
100	10	250	250.000	
20	10	500	100.000	
0	20	-500		
20	20	-250		100.000
15	20	0		
70	20	250	350.000	
95	20	500	950.000	
		T O T A L :	2.600.000	-337.000

Deci valoarea sumei duble de la numărător va fi:

$$2.600.000 - 337.500 = 2.262.500$$

Urmează apoi că:

$$n.d_x d_y = 93.750$$

şi deci valoarea numărătorului va fi:

$$2.262.500 - 93.750 = 2.168.750$$

Calculele de la numitor ne conduc la:

$$n . \sigma_y . \sigma_y = 1000.12,2.298,7 = 3.644.140$$

şi deci

$$r = \frac{2.168.750}{3.644.140} = 0,60$$

Să prezentăm în continuare cîteva proprietăţi ale coeficientului de corelaţie, însoţite şi de unele precizări.

a) Aşa cum am spus, intervalul de variaţie pentru r este de la -1 la +1. In cazul în care se ating aceste valori extreme, corelaţia poate fi numită perfectă (pozitivă sau negativă). Valoarea nulă a coeficientului indică o situaţie în care cele două variabile sînt necorelate.

b) O condiţie suficientă pentru ca r să atingă valoarea 1 este ca cele două variabile să ia aceleaşi valori, ca distribuţiile lor de frecvenţă să coincidă (adică frecvenţele marginale ale tabelului de corelaţie să fie identice) şi ca toate frecvenţele interioare ale tabelului să se afle pe diagonala principală. Sau, schematic:

$$\left.\begin{array}{l} s = t \\ x_i = y_i \\ k_{.i} = k_{i.} \\ k_{ii} = k_{.i} \end{array}\right\} \Longrightarrow r = 1$$

Proprietatea reciprocă nu este întotdeauna adevărată, în sensul că există situaţii în care valorile x_i şi y_i nu coincid,

dar coeficientul ia valoarea 1. In orice caz, corelaţia perfectă nu se poate atinge decît dacă tabelele sînt pătratice (s=t) şi toate frecvenţele interioare sînt diagonale.

c) Dacă frecvenţele interioare, k_{ij}, verifică relaţia de independenţă:

$$k_{ij} = \frac{k_{i.} k_{.j}}{n}$$

atunci valoarea lui r este nulă. Reciproca nu este nici aici adevărată, putîndu-se obţine valori nule pentru r şi cînd frecvenţele nu verifică relaţia de independenţă. Prin urmare, independenţa implică necorelaţea variabilelor sau, invers, existenţa corelaţiei implică asocierea (în sensul definit la caracteristicile calitative). Din faptul că afirmaţiile reciproce nu sînt adevărate, rezultă cele ce spuneam la început, şi anume că cele două noţiuni aplicabile în cazul variabilelor (asocierea şi corelaţia) nu se suprapun exact.

d) O altă proprietate remarcabilă a coeficientului de corelaţie constă în faptul că valoarea lui nu este, la drept vorbind, o funcţie de valorile variabilelor, ci de diferenţele dintre aceste valori, luate două cîte două, deci de distanţele dintre indivizi şi mai exact de rapoartele între aceste distanţe. Asta înseamnă că putem deplasa spre stînga sau spre dreapta seria de valori şi, de asemenea, că putem înmulţi sau împărţi cu un acelaşi număr valorile unei variabile fără a se modifica valoarea lui r. Sau, altfel spus, coeficientul de corelaţie rămîne constant dacă efectuăm transformări lineare ale celor două variabile:

$$X' = a + bX$$
$$Y' = c + dY$$

Importanţa proprietăţii de faţă nu constă numai în faptul că prin asemenea transformări putem uşura în mod simţitor calculele, destul de laborioase altfel, ce intervin în formule (1), (2) şi (3). Ea devine evidentă pentru situaţiile în care procedura de atribuire a valorilor nu este cu totul riguroasă, acesta fiind un caracter aproape constant al variabilelor

sociale. Stim astfel că, pentru a ajunge la un "adevărat" coeficinet de corelaţie între două variabile, nu este necesar să măsurăm exact valorile lor; va fi suficient să găsim o procedură ce permite atribuirea valorilor şi care să reproducă în mod cosect rect raportul intervalelor. Dacă notăm cu 1,2,...,s valorile caracteristicii, important, pentru analiza de corelaţie, nu este faptul că indivizii iau sau nu aceste valori şi nici că diferenţa dintre valoarea unei clase şi cea a precedentei este unitatea, ci numai că distanţele între indivizii aparţinînd la două clase consecutive sînt constante.

Revenind la problematica simplificării calculelor, menţionez un caz destul de frecvent întîlnit, acela în care valorile uneia sau ambelor variabile sînt în progresie aritmetică. Atunci aceste valori se pot transforma în mulţimea numerelor naturale sau întregi, păstrînd caracterul de progresie aritmetică. De exemplu, în exerciţiul prezentat mai sus, atît pentru valorile vîrstei cît şi pentru cele ale retribuţiei putem folosi numere întregi în valoare absolută mult mai mică. Cea mai importantă simplificare se obţine atunci cînd o valoare mijlocie se transformă în zero iar celelalte în ± 1, ±2 . In exemplul nostru putem lua ca valori pentru ambele variabile următoarele: -2, -1, 0, 1, şi 2. urmînd apoi să aplicăm direct formula (2) în calculul lui <u>r</u>. Să reluăm acest exemplu:

<u>Pentru X</u> :

x_i	k_i	$k_i . x_i$	x^2	$k_i . x_i^2$
-2	1oo	-2oo	4	4o1
-1	15o	-15o	1	15o
0	35o	-	-	-
1	2oo	2oo	1	2oo
2	2oo	4oo	4	8oo
Total	1ooo	25o		155o

Luînd pe 0 valoare ajutătoare, rezultă:

$\bar{x} = 0,25$ $d_x = 0,25$ $d_x^2 = 0,0625$

$s_x^2 = 1,55$ $\sigma_x^2 = 1,4875$ $\sigma_x = 1,22$

<u>Pentru Y:</u>

y_j	$k_{.j}$	$k_{.j}y_j$	y_j^2	$k_{.j}y_j^2$
-2	100	-200	4	400
1	200	-200	1	200
0	300	-	-	-
1	250	250	1	250
2	150	300	4	600
Total	1000	150		1450

Si

$\bar{y} = 0,15$ $d_y = 0,15$ $d_y^2 = 0,0225$

$s_y^2 = 1,45$ $\sigma_y^2 = 1,4275$ $\sigma_y = 1,19$

Calculul sumei duble este prezentat mai jos, fără a mai reproduce cazurile în care produsele sînt nule.

k_{ij}	x_i	y_j	$k_{ij}x_iy_j$ +	−
50	-2	-2	200	
30	-2	-1	60	
5	-2	1		10
25	-1	-2	50	
70	-1	-1	70	
15	-1	1		15
10	-1	2		20
10	1	-2		20
30	1	-1		30
100	1	1	100	
20	1	2	40	

2o	2	-1		4o	
7o	2	1	14o		
95	2	2	38o		
TOTAL:			104o	-135	= 9o5

Deci suma dublă este 9o5 şi urmează să calculăm:

$$n \cdot \bar{x} \cdot \bar{y} = 1ooo.o,25.o,15 = 37,5$$

ceea ce înseamnă că numărătorul coeficientului va fi:

$$9o5 - 37,5 = 867,5$$

Pentru numitor, înmulţim

$$n.\sigma_x \sigma_y = 1ooo.1,22.1,19 = 1451,8$$

şi deci

$$r = \frac{867,5}{1451,8} = o,6o$$

adică s-a ajuns la acelaşi rezultat cu un volum mai mic de calcule.

Desigur, că în condiţiile în care cercetătorul dispune de un calculator electronic pentru prelucrarea datelor, toată această poveste despre simplificarea operaţiilor nu-şi mai are rostul. Am ţinut totuşi să ofer procedeele respective pentru motivul că încă multe calcule se mai efectuează manual. Procedăm aşa mai ales atunci cînd calcularea unui coeficient de corelaţie (sau a unui alt indicator) este o problemă în sine şi imediată, deci cînd nu este inclusă într-un program mai amplu de prelucrare a datelor, pentru care utilizarea calculatorului să fie dacă nu indispensabilă cel puţin foarte eficientă. Anumite dificultăţi legate de accesul la maşinile electronice de calcul, lipsa de pregătire a majorităţii oamenilor pentru a purta un dialog eficient cu computerul şi altele fac însă rentabile unele calcule cu maşini de birou, operaţii ce nu depăşesc un volum de muncă de cîteva ore. Or, aflarea unuia sau cîtorva coeficienţi de corelaţie, chiar pentru variabile mai complexe decît cele din exemplul dat, este o muncă ce se încadrează în acest cuantum de timp.

Un caz particular. Dacă problema asocierii se pune doar atunci cînd avem observații asupra unui număr mare de indivizi statistici - numai în situații cu totul deosebite, un tabel de asociere, chiar și pentru caracteristici dihotomice, spune ceva dacă cuprinde mai puțin de 1oo de cazuri - corelația poate fi urmărită iar coeficientul de corelație calculat, și pentru loturi mici de indivizi. In asemenea situații, nu mai are rost construcția caracteristicilor cu intervale de grupare, deoarece este mai comodă înșiruirea indivizilor într-o anumită ordine și evidențierea valorilor pe care le ia fiecare în raport cu X și Y. Atunci, în formulele care dau coeficientului, vor dispare frecvențele k_{ij}, însumarea se va face urmărindu-se indivizii așezați într-o anumită ordine și, prin urmare, suma dublă se va transforma în una simplă, după un singur indice, mergînd de la 1 la n. Astfel, formula (1) va deveni:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \sigma_y} \qquad (4)$$

Si în mod similar se pot scrie și celelalte două.

Să luăm pentru exemplificare o situație foarte simplă. Să presupunem că pentru 1o județe - pe care le-am simbolizat cu litere de la A la J - dorim să vedem ce corelație există între numărul cinematografelor care funcționează în fiecare județ (X) și numărul spectatorilor de film într-un an (Y). Cifrele sînt și de această dată fictive deoarece pentru scopuri didactice sînt preferabile exemple special construite.

Județul	A	B	C	D	E	F	G	H	I	J	Total
Nr. de cinematografe(X)	1oo	12o	13o	14o	15o	16o	17o	19o	21o	23o	16oo
Nr. spectatori, în milioane(Y)	3,1	2,4	6,4	2,8	5,9	4,0	5,6	5,0	6,8	8,0	50

Tabel 3. Numărul cinematografelor și spectatorilor de film în 1o județe (cifre fictive)

Calculele se vor efectua în felul următor:

Jud.	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i-\bar{y})^2$	$(x_i - \bar{x})(y_j- \bar{y})$	
A	100	3,1	-60	-1,9	3600	3,61	114	
B	120	2,4	-40	-2,6	1600	6,76	104	
C	130	6,4	-30	1,4	900	1,96		42
D	140	2,8	-20	-2,2	400	4,84	44	
E	150	5,9	-10	0,9	100	0,81		9
F	160	4,0	0	-1,0	0	1,00		
G	170	5,6	10	0,6	100	0,36	6	
H	190	5,0	30	0	900	0		
I	210	6,8	50	1,8	2500	3,24	90	
J	230	8,0	70	3,0	4900	9,00	210	
Tot.	1600	50,0			15000	31,58	568 - 51 = 517	

Din primele două sume, rezultă evident că:

$$\bar{x} = 160; \qquad \bar{y} = 5$$

valori cu ajutorul cărora s-au calculat expresiile din coloanele următoare. (Dacă n-ar fi ieşit cifre rotunde, se puteau folosi valori ajutoare ca şi mai înainte). Urmează apoi că:

$$\sigma_x = \sqrt{1500} = 38,73$$

$$\sigma_y = \sqrt{3,158} = 1,78$$

In fine, ultima coloană dă numărătorul coeficientului de corelaţie:

$$r = \frac{517}{10.38,73.1,78} = 0,75$$

Am obţinut deci o valoare pozitivă ridicată, ce ilustrează o legătură strînsă între numărul cinematografelor şi cel al spectatorilor, în judeţele analizate. Urmărind ultima coloană, observăm că doar două din cele zece produse sînt negative (pentru judeţele C şi E), avînd de a face aici cu ceea ce numeam la început discrepanţă de comportament, în ambele cazuri numărul de

cinematografe fiind sub media lotului analizat iar numărul spectatorilor peste medie.

Doresc să atrag atenţia că foarte des în texte - aşa cum am făcut şi eu mai sus - mărimea coeficienţilor de corelaţie este apreciată calitativ: " valoare ridicată", "coeficient foarte mare", corelaţie slabă" etc. Menţionez că nu există, nişte limite stabilite între care să se găsească valoarea coeficientului pentru a i se atribui asemenea etichete calitative. Aprecierile depind de cele mai multe ori de aşteptările cercetătorului, iar nu o dată se întîmplă ca un acelaşi coeficient să fie apreciat de unii drept mare şi de alţii mic. Chiar şi în exemplul anterior, avînd de a face cu două variabile care în mod "normal" trebuie să fie corelate, valoarea o,75 poate fi contradictoriu apreciată, după cum este sub sau peste aşteptările celui care efectuează analiza.

In sfîrşit, mai menţionez că adesea se vorbeşte despre valori "semnificative" sau "nesemnificative" ale lui r. Se cuvine subliniat că termenul de "semnificativ" are în statistică o semnificaţie foarte clară, referindu-se exclusiv la mărimile obţinute asupra eşantioanele probabiliste. De pildă, în exemplul precedent, dacă lo judeţe constituie obiectul exclusiv al studiului nostru, coeficientul obţinut (chiar dacă ar fi fost numai o,o5 să zicem) nu este nici semnificativ nici nesemnificativ; el este atît cît este. Dacă însă, cele lo judeţe ar fi fost considerate un eşantion din toate judeţele ţării, atunci problema semnificaţiei lui r s-ar pune, avînd menirea să dea răspuns la întrebarea dacă nu cumva, prin selecţie, s-a obţinut din întîmplare o valoare diferită de zero. Există proceduri statistice bine elaborate pentru a soluţiona o asemenea chestiune.

2. Corelaţia rangurilor

Problema corelaţiei rangurilor este, în aparenţă, diferită prin conţinut de cele discutate aici, locul său părînd mai degrabă a fi în capitolul ce tratează asocierea. In esenţă, e vorba despre următorul lucru. Să presupunem că avem un grup relativ mic de indivizi şi două criterii de ordonare (clasificare ierarhică) a acestora. Se pune problema de a găsi un coeficient care să exprime gradul de similitudine al ordonării după cele două criterii.

Putem să transcriem aceasta într-un limbaj mai apropiat de cel utilizat în cursul de față. Ordonarea indivizilor înseamnă , de fapt, aplicarea unei <u>scale ordinale</u> pentru o caracteristică dată. In cazul unei ordonări <u>stricte</u>, fiecare treaptă a scalei va conține numai un singur individ din populație, deci scala are atîtea trepte cîți indivizi sînt. Să presupunem, momentan, că ne găsim într-o asemenea situație simplă. Atunci apare cu claritate că problema noastră este acea a calculării unui coeficient de corelație între două caracteristici <u>calitative</u>, fapt care intră în contradicție cu precizările anterioare, și anume că vom vorbi de corelație doar în cazul caracteristicilor cantitative. Vom vedea imediat cum se înlătură acest paradox.

Să începem prin a clarifica mai mult datele problemei. Fie o mulțime de <u>n</u> indivizi statistici pe care să-i botezăm cu:

$$A_1 , A_2 , \ldots\ldots , A_n$$

unde indicii 1,2,...,n nu sînt decît niște etichete pentru identificare și pentru a arăta că în total sînt <u>n</u> indivizi. Ordinea scrierii lor nu are, așadar, nici o importanță. Să presupunem că, în raport cu un criteriu X, am reușit să realizăm o ordonare strictă a indivizilor, de la cel care posedă proppietatea în cea mai mare măsură la cel care o posedă în cea mai mică, de pildă. Aceasta nu înseamnă altceva decît a stabili o corespondență biunivocă între mulțimea indivizilor și o mulțime $\{x_1, x_2, \ldots, x_n\}$ formată din numere naturale diferite, cuprinse între 1 și <u>n</u> (o permutare a numerelor 1,2,...,n) Astfel, dacă lui A_1 îi vom pune în corespondență numărul x_1 egal cu 3 înseamnă că individul A_1 se află pe locul 3 al clasamentului stabilit după X ș.a.m.d. Procedînd similar cu Y, obținem un alt clasament, respectiv o altă permutare a aceleiași mulțimi de numere: $\{y_1, y_2, \ldots, y_n\}$. Finalmente vom ajunge la un tabel de forma:

Indivizi	A_1	A_2	...	A_n
Rang după X	x_1	x_2	...	x_n
Rang după Y	y_1	y_2	...	y_n

Pentru a înţelege perfect despre ce este vorba, că dăm un exemplu foarte la îndemînă. Fie mulţimea indivizilor formată din cele 18 echipe de fotbal care participă într-un an la campionatul primei divizii şi fie $x_1, x_2, \ldots, x_{18}$ locul pe care fiecare echipă îl ocupă în clasament la sfîrşitul turului iar $y_1, y_2, \ldots, y_{18}$ locurile ocupate de aceleşi echipe la terminarea campionatului. Un coeficient de corelaţie a rangurilor are, cum spuneam, rolul să indice similitudinea, apropierea dintre cele două clasamente. O similitudine perfectă - adică, în cazul nostru, menţinerea de către fiecare echipă a locului în clasament - va trebui să ne conducă la o valoare +1 a coeficientului, iar o discrepanţă totală - adică inversarea clasamentului în varianta a doua - , la o valoarea -1. Mărimea nulă va trebui să se atingă atunci cînd între cele două clasamente nu există nici o legătură, respectiv cînd unul dintre ele dispune aleator indivizii în raport cu celălalt.

Mărirea care satisface aceste cerinţe se notează cu ς şi se numeşte coeficientul de corelaţie a rangurilor sau coeficinetul lui Spearman. Formula sa este:

$$\varsigma = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n^3 - n}$$

unde cu d_i am notat diferenţele rangurilor (poziţiilor) fiecărui individ în cele două clasamente:

$$d_i = x_i - y_i \quad ; \; i = 1,2,\ldots,n$$

Practic, calculul lui ς este foarte simplu. Datele se trec într-un tabel ca cel de mai sus, doar că se scriu pe verticală (pentru a se facilita însumarea) şi se mai adaugă două coloane: una pentru valorile d_i, alta pentru d_i^2.

Să reluăm exemplul cu judeţele şi în loc să folosim efectiv numărul cinematografelor şi cel al spectatorilor să ierarhizăm judeţele după aceste două criterii (de fapt, se vede că în raport cu prima variabilă judeţele au şi fost scrise în ordine) Vom ajunge la situaţia din tabelul 4.

Județul	Rang după X	Rang după Y	d_i	d_i^2
A	1o	8	2	4
B	9	1o	-1	1
C	8	3	5	25
D	7	9	-2	4
E	6	4	2	4
F	5	7	-2	4
G	4	5	-1	1
H	3	6	-3	9
I	2	2	0	0
J	1	1	0	0
TOTAL				52

Tabelul 4. Ierarhia județelor după variabilele din tabelul 3.

Deci suma pătratelor diferențelor este 52, și aplicînd formula:

$$\varrho = 1 - \frac{6 \cdot 52}{1000-10} = 1 - \frac{312}{990} = 1 - 0{,}32$$

$$\varrho = 0{,}68$$

După efectuarea acestor calcule intuim mai ușor că ϱ satisface proprietățile cerute. Dacă cele două clasamente sînt identice, atunci toate diferențele vor fi nule și deci tot nul va fi și termenul care se scade din 1. Cu cît cele două clasamente sînt mai deosebite unul de celălalt, cu atît diferențele vor fi mai mari și deci termenul care se scade mai mare. La atingerea de către acesta a valorii 1 (ϱ = 0), clasamentele sînt independente, iar la atingerea lui 2(ϱ = -1) clasamentele se inversează.

Să mai luăm un exemplu. Să presupunem că doi profesori (de pildă, unul de muzică şi altul de educaţie fizică) au fost solicitaţi să ordoneze 15 copii în funcţie de aptitudinile la disciplinele respective. Clasamentele întocmite de către cei doi profesori (imaginari) apar în tabelul 5, ca şi restul elementelor pentru calcularea lui ρ.

Din datele tabelului 5, se obţine:

$$\rho = 1 - \frac{6.560}{3360} = 1 - \frac{3360}{3360} = 1 - 1 = 0$$

Deci se poate afirma că, în lotul studiat, aptitudinile la cele două obiecte sînt aleator distribuite, adică nu se manifestă nici o tendinţă ca cei care sînt mai buni la muzică să fie mai buni ori mai slabi la sport.

Elevii	Clasament la muzică (X)	Clasament la sport (Y)	d_i	d_i^2
A	7	10	-3	9
B	3	11	-8	64
C	1	5	-4	16
D	4	8	-4	16
E	1o	2	8	64
F	15	6	8	81
G	12	15	-3	9
H	11	1	1o	1oo
I	8	3	5	25
J	2	9	-7	49
K	6	4	2	4
L	5	12	-7	49
M	9	14	-5	25
N	13	13	0	0
O	14	7	7	49
TOTAL:				56o

Tabelul 5. Clasamentele celor 15 elevi, după aptitudinile la două discipline

Coeficientul de corelaţie a rangurilor este foarte de utilizat în ştiinţele umane (mai ales în psihologie, unde prin tradiţie se lucrează cu eşantioane mici) şi nu o dată cei care scriu despre el îi laudă virtuţile . Fascinaţia pe care o exercită acest coeficient se datorează iluziei că el ar reprezenta o măsură a unei corelaţii între caracteristici calitative, car deci nu implică operaţia de măsurare. Se poate arăta uşor că această iluzie se spulberă imediat dacă ne dăm osteneala să înţelegem de unde provine formula lui Spearman.

Nu este locul să prezint aici deducaţia matematică a acestei formule. Este suficient să precizez că, de fapt, coeficientul ρ nu este decît un obişnuit coeficient de corelaţie r, obţinut prin aplicarea formulei (4) a lui r, în care valorile x_i şi y_i sînt numere naturale distincte între 1 şi n. Datorită faptului că în toate situaţiile valorile lui X şi Y sînt mereu cele amintite mai sus şi că fiecare valoare este luată o singur dată este clar că se va ajunge la o formulă de simplificare a coeficientului. De pildă, mediile lui X şi Y vor fi egale şi constante, pentru n dat, indiferent de clasamente, căci:

$$\bar{x} = \bar{y} = \frac{1+2+\ldots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}$$

La fel vor fi egale şi constante şi abaterile standard iar suma de la numărătorul coeficientului se va exprima în funcţie de n şi de diferenţele d_i.

Esenţial este deci faptul că ρ este tot un coeficient de corelaţie, făcînd în consecinţă, apel tot la valori ale variabilelor. Scamatoria ni se dezvăluie atunci imediat: s-a utilizat pentru fiecare variabilă o clasificare; indivizii au primit un număr care le desemna locul în clasament iar apoi acest număr a fost luat drept valoare. S-a efectuat deci o trecere ta cită de la numere ordinale la numere cardinale.

Gravitatea "infracţiunii" este substanţial diminuată de proprietatea coeficientului de corelaţie menţionată la punctul

d), in paragraful precedent. Intr-adevăr, atribuirea valorilor 1,2,....,n nu înseamnă - în calculul lui ρ - că indivizii iau efectiv aceste valori, ci numai că distanțele dintre ei sînt egale. Drept urmare, coeficientul de corelație a rangurilor este efectiv un coeficient de corelație, aplicat atunci cînd nu este posibilă operația de măsurare dar se pot ierarhiza indivizii studiați, presupunînd că diferențele de despart doi indivizi consecutivi sînt identice.

Se înțelege atunci că, ori de cîte ori este posibil, trebuie preferat calculul lui r și nu cel al lui ρ (lucrul fiind evident, de pildă, în exemplul cu județele). Cerința unor metodologi de a se folosi corelația rangurilor atunci cînd măsurătorile sînt imprecise nu este totodeauna justificată. Să presupunem că dorim să corelăm două variabile cum ar fi pregătirea unor elevi la două obiecte de învățămînt. Elevii avînd note la respectivele materii corelația se poate calcula folosind notele în calitate de valori sau cu ajutorul coeficientului lui Spearman, după clasamentele ce se pot întocmi pe baza acelorași note. Unii pedagogici preferă aceasta a doua cale pentru motivul că notele nu reprezintă o măsură exactă a nivelului de cunoștințe al elevilor. Ei nu înțeleg că aceeași imprecizie poate duce la o clasificare greșită. Mai mult, să zicem că la un obiect elevii L,M și N vor avea notele 8, 7,5o și respectiv, 7,45 apărînd în această ordine în clasament. Cele 5 sutimi de punct ce-l despart pe N de M pot fi datorate unei întîmplări și este foarte greu să acceptăm că distanța dintre L și M va fi egală cu cea dintre M și N. Or, în calculul coeficienților de corelație distanțele contează în primul rînd.

Ranguri întrunite. Se știe că uneori clasamentele nu se pot întocmi în mod perfect, adică 2,3 sau chiar mai mulți indivizi se pot clasa pe același loc. Atunci, strict vorbind, formula scrisă pentru ρ nu se mai poate aplica, existînd elaborată o alta care ține seama de numărul rangurilor întrunite sau legate. Cum ea este destul de complicată, nu o voi mai prezenta aici, precizînd doar că, dacă numărul rangurilor întrunite este mic, se poate folosi și formula lui Spearman, acceptînd că aceasta ne dă o aproximație a c[illegible]ientului de corelație adevărat.

In astfel de cazuri, trebuie să acordăm atenţie felului cum se notează rangurile. Să presupunem că, într-un clasament, după individul care ocupă primul loc urmează doi la egalitate. Ei vor corespunde poziţiilor 2 şi 3, dar cum nu-i putem diferenţia trebuie să le acordăm aceeaşi valoare, care va fi, în cazul imaginat, 2.5 iar individului care urmează după cei doi îi vom acorda rangul 4. Deci dacă mai mulţi indivizi ocupă aceeaşi poziţie în clasament, fiecare va primi o valoare egală cu rangul mediu, pe cînd individul care urmează primeşte acea valoare ce i-ar fi revenit dacă precedenţii ar fi fost diferenţiaţi.

3. Regresie, corelaţie parţială şi multiplă

Şi în cazul variabilelor statistice există proceduri - şi încă mult mai bine puse la punct decît pentru caracteristicile calitative - de analiză a raporturilor simultane dintre mai mulţi factori. Este evident o problemă de generalizare a coeficientului de corelaţie la situaţia mai complexă cînd numărul variabilelor studiate depăşeşte pe 2. Pentru a putea înţelege semnificaţia coeficienţilor generalizaţi, chiar şi în maniera foarte sumară în care îi vom prezenta în paragraful de faţă, va trebui să specificăm şi alte interpretări ale coeficientului simplu de corelaţie, interpretări ce se desprind din analiza noţiunii de <u>regresie</u>. Ar fi greşit însă să se înţeleagă că regresia este doar o procedură statistică ce ne ajută să facem trecerea amintită; dimpotrivă, analiza de regresie reprezintă un capitol fundamental al statisticii parametrice, important în sine, important şi în elucidarea raportului dintre două variabile, lucru care cred că se va înţelege din rîndurile ce urmează.

<u>Regresia în cazul a două variabile.</u> Fie X şi Y două variabile statistice, definite pe o mulţime de <u>n</u> indivizi. Pentru a facilita înţelegerea rîndurilor ce urmează, să presupunem că acest număr este mic, deci nu efectuăm restrîngeri pe intervale de grupare şi evidenţiem fiecare individ prin valorile x_i şi y_i pe care le ia, supoziţie prin care nu se restrînge deloc generalitatea problemei.

Putem da fenomenului o ilustrare grafică foarte simplă.

Dacă luăm în plan două axe de coordonate ortogonale şi marcăm pe una valorile în raport cu X şi pe cealaltă în raport cu Y, atunci fiecare individ primeşte ca imagine un punct din plan de coordonate (x_i, y_i). Imaginea de ansamblu a tuturor celor n indivizi studiaţi se va prezenta ca un "nor" de asemenea puncte.

Se poate uşor înţelege că forma acestui "nor" ne va da o indicaţie asupra existenţei şi sensului unei legături de corelaţie între cele două variabile, după tendinţe de cuplare a valorilor lui X cu cele ale lui Y. Aşa de pildă, în cazul unei corelaţii pozitive, cînd deci valorile mari ale lui X se leagă, în medie, cu valori mari ale lui Y şi cele mici cu cele similare, norul de puncte va fi alungit pe o direcţie apropiată de cea a bisectoarei întîi a sistemului de axe (cea care trece prin cadranele I şi III) iar în caz contrar, se va contura un nor alungit după o direcţie similară cu cea a bisectoarei a doua. Cele două cazuri sînt redate în figurile 1 şi 2.

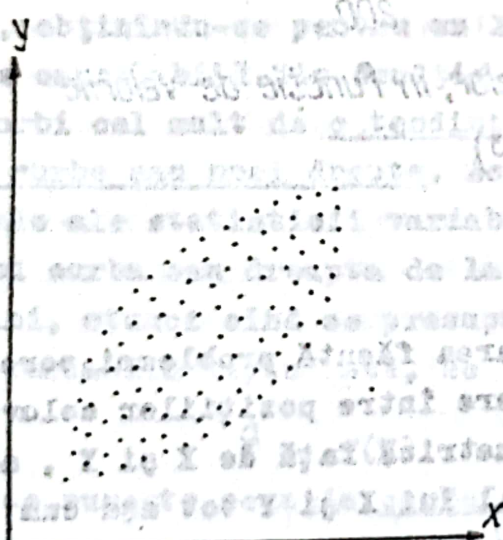


Fig.1 Nor de puncte sugerînd o corelaţie pozitivă

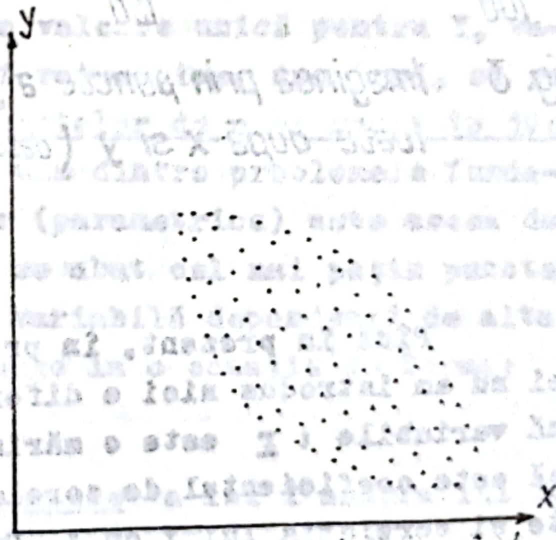


Fig.2 Nor de puncte sugerînd o corelaţie negativă

Ca să prezentăm o situaţie şi mai concretă, să reluăm exemplul cu judeţele utilizat deja în acest capitol. Reprezentînd grafic situaţia din tabelul 3, vom ajunge la imaginea din figura 3.

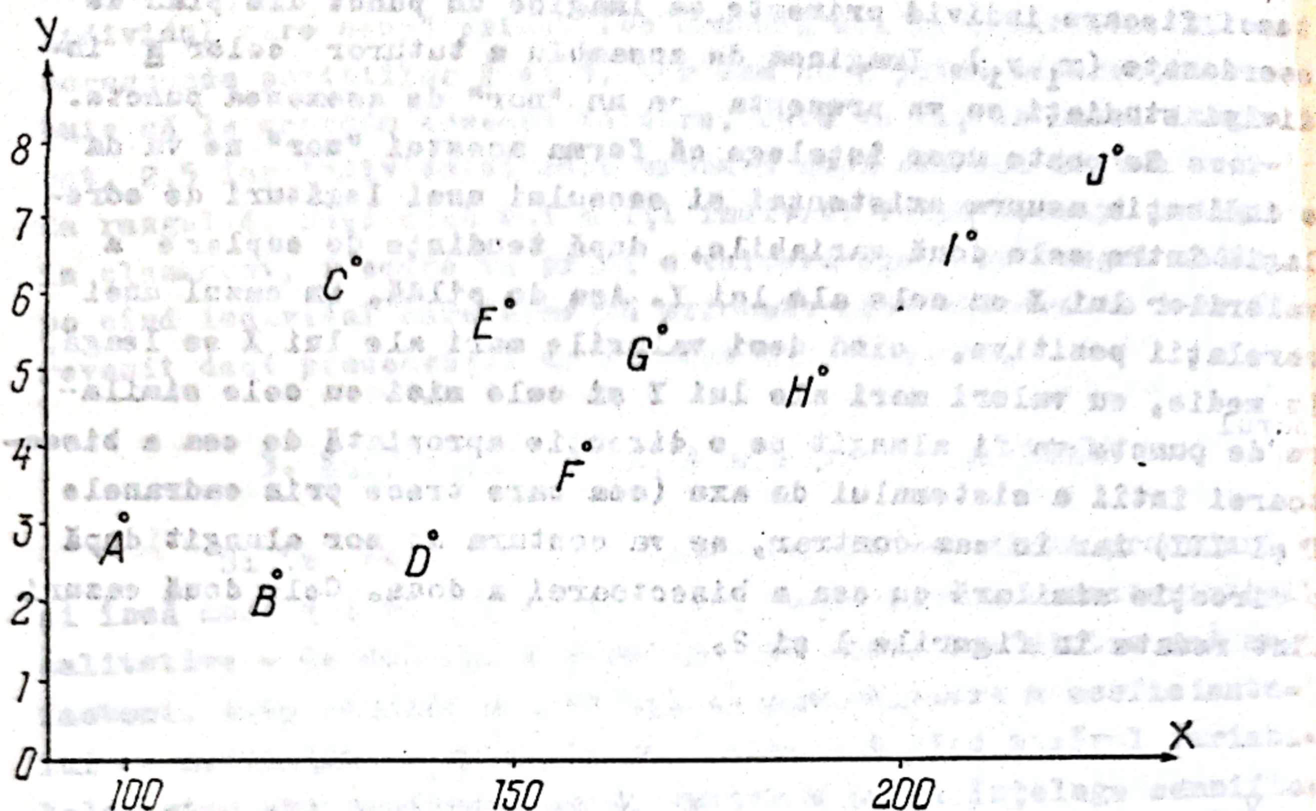


Fig. 3 Imaginea prin puncte a județelor, în funcție de valorile luate după x și y (datele tab. 3)

Pînă în prezent, în prezentarea făcută problemei corelației nu am introdus nici o diferențiere între pozițiilor celor două variabile : <u>r</u> este o mărime simetrică față de X și Y , adică este coeficientul de corelație al lui X și Y tot așa cum este și corelația lui Y cu X. De acum încolo lucrurile se vor schimba puțin, căci vom presupune una dintre variabile ca fiind dependentă de cealaltă, în sensul matematic al cuvîntului, adică vom zice că variația uneia antrenează variația celeilalte. Să nu uităm încă că ne aflăm în domeniul statisticii. Dacă dependența ar fi funcțională (deci riguroasă), am zice că unei creșteri a lui X cu o valoare dată îi corespunde o creștere sau o descreștere a lui Y cu altă valoare dată. In dependențele de tip statistic, formularea poate îmbrăca învelișul verbal următor: unei creșteri a lui X cu o valoare dată îi corespunde <u>în medie</u> o creștere

a lui Y cu o altă valoare.

Obsevație. Aceste dependențe formale, indiferent că sînt funcționale sau statistice, nu reflectă în mod obligatoriu dependențe materiale între fenomenele pe care variabilele le reprezintă. Se poate ca, de pildă, unei creșteri a numărului de cinematografe (X) să-i corespundă - așa cum s-a întîmplat în exemplul nostru - o creștere, în medie, a numărului de spectatori (Y). Relația efectivă poate fi în ambele sensuri: fie că numărul sporit al cinematografelor constituie o condiție favorizantă pentru atragerea unui număr mai mare de spectatori, fie că volumul acestor instituții este determinat de efectivul demografic al județelor, deci de numărul potențial de spectatori. Probabil că aceasta a doua alternativă este mai plauzibilă, dar asta nu ne împiedică să-l calculăm pe "Y în funcție de X".

Intuitiv, dependența statistică dintre variabile este redată de forma norului de puncte; dacă dependența ar fi funcțională, atunci punctele s-ar plasa exact pe o dreaptă sau pe o curbă oarecare, obținîndu-se pentru un X dat o valoare unică pentru Y, valoare calculabilă din ecuația curbei respective. Așa însă, se poate vorbi cel mult de o tendință a punctelor de a se grupa în jurul unei curbe sau unei drepte. Astfel, una dintre problemele fundamentale ale statisticii variabilelor (parametrice) este aceea de a găsi curba sau dreapta de la care se abat cel mai puțin punctele-imagini, atunci cînd se presupune o variabilă dependentă de alta. Cu Y variabilă dependentă, se va ajunge la o ecuație de forma:

$$\hat{Y} = f(X)$$

care se numește ecuația curbei de regresie a lui Y asupra lui X. In particular, cînd

$$\hat{Y} = a + bX \qquad (1)$$

se va obține dreapta de regresie.

In continuare mă voi mărgini la cîteva considerații asupra regresiei lineare, datorită faptului că este cazul cel mai simplu și pentru că, în realitate, este și cel mai frecvent întîlnit.

Așadar, avînd mulțimea de puncte reprezentînd indivizii noștri statistici, problema este de a determina o dreaptă de forma (1) de așa manieră încît ca să se afle cît mai aproape de totali-

tatea acestor puncte. Faptul că este important care dintre cele două variabile este socotită independentă şi care depdentă va reieşi imediat cînd se va preciza sensul expresiilor "cea mai mică abatere a punctelor de la dreaptă" sau "cea mai apropiată dreaptă de puncte".

Mai întîi, trebuie să înţelegem că determinarea dreptei de regresie ne pune în posesia unei dependenţe funcţionale între două mărimi X şi Y. Cum însă dependenţa dintre variabilele noastre iniţiale nu este, în general, de această formă, devine limpede că variabila $\hat{Y}$ din ecuaţia (1) nu este nici ea variabila Y iniţială, ci una nouă ale cărei valori rezultă efectuînd calcule în membrul drept al relaţiei (1), ceea ce explică deci notaţia folosită. Relaţia (1) exprimă tendinţa medie de covariaţie a lui X şi Y şi ea nu are drept scop să prezică nişte coordonate individuale, ci un comportament mijlociu, de grup. Intr-adevăr, dacă x_i este valoarea pe care un individ i o ia în raport cu X, este foarte probabil ca valoarea y_i pe care acelaşi individ o ia faţă de Y să nu coincidă cu valoarea $\hat{y}_i$ calculată cu formula (1) - înmulţind pe x_i cu b şi adunînd apoi a - datorită faptului că dreapta de regresie nu poate trece prin toate punctele decît în cazurile de excepţie cînd acestea sînt colineare. Ba mai mult, un alt individ j poate lua după X aceeaşi valoare ca şi precedentul ($x_i = x_j$) dar după Y o alta ($y_i \neq y_j$). Or, dreapta de regresie, ca orice dreaptă, realizează o corespondenţă biunivocă între valorile variabilelor, adică la o abscisă corespunde o singură ordonată. Prin urmare, dacă ne plasăm în cazul cînd prin indicele i desemnăm indivizii, înseamnă că pentru orice individ din populaţie vom avea un triplet de valori:

x_i - valoarea în raport cu X;

y_i - valoarea în raport cu Y;

$\hat{y}_i$ - valoarea calculată din ecuaţia dreptei de regresie

Cum, în general, ultimele două valori nu coincid; diferenţele dintre ele de forma

$$y_i - \hat{y}_i$$

redaţi mărimea abaterilor punctelor de la dreapta de regresie, respectiv eroarea care se introduce atunci cînd în loc de valoarea efectivă y_i lucrăm cu valoarea calculată $\hat{y}_j$. Atunci este evident că dreapta căutată de noi va fi cea care, pe ansamblu, va realiza cele mai mici abateri de genul precizat. Acest "pe ansamblu" nu înseamnă simplu suma diferenţelor, deoarece ele fiind şi pozitive şi negative prin adunare se pot anula chiar la valori absolute foarte mari. Va trebui deci înlăturat întîi semnul diferenţelor ceea ce se realizează prin ridicarea lor pătrat, şi abia apoi se trece la însumare. Se obţine deci expresia:

$$E_y = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

sau, cu expresia lui $\hat{y}_i$ din (1)

$$E_y = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \qquad (2)$$

care reprezintă suma pătratelor abaterilor valorilor calculate de la valorile observate, pentru variabila Y.

Coeficienţii a şi b ai dreptei de regresie se determină de aşa manieră încît mărimea E_y să fie minimă. Determinarea acestor coeficienţi este întotdeauna posibilă şi procedura- numită curent în literatura statistică metoda celor mai mici pătrate - conduce la o soluţie unică, ceea ce vrea să spună că, oricare ar fi forma norului de puncte, există întotdeauna o dreaptă de forma (1) care satisface condiţia impusă. Deci dreapta de regresie există şi este unică.

Trecînd peste calculele algebrice care conduc la determinarea coeficienţilor a şi b, voi menţiona doar formulele la care se ajunge:

$$b = r \frac{\sigma_y}{\sigma_x} \qquad (3)$$

$$a = \bar{y} - r \frac{\sigma_y}{\sigma_x} \bar{x} \qquad (4)$$

unde simbolurile din dreapta au semnificaţia cunoscută: coeficientul de corelaţie dintre X şi Y (r), abaterile standard şi valorile medii ale celor două variabile.

Să vedem acum care este interpretarea coeficienţilor din ecuaţia dreptei. Coeficientul a ne dă intersecţia dreptei cu axa OY şi are mai puţină importanţă pentru cele ce urmează. Valoarea b, care se numeşte panta sau coeficientul unghiular al dreptei, arată cu cîte unităţi se schimbă valoarea lui Y atunci cînd X creşte cu o unitate (e vorba de unităţile de măsură specifice fiecărei variabile). De aici ajungem şi la interpretarea coeficientului de corelaţie care intervine în expresia lui b. Să presupunem că variem pe X nu cu o unitate, ci cu o abatere standard, σ_x, ceea ce va produce pentru $\hat{Y}$ o variaţie Δy, pe care urmează să o calculăm. Adică:

$$\hat{Y} + \Delta y = a + b(X + \sigma_x)$$

sau

$$\hat{Y} + \Delta y = a + bx + b\,\sigma_x$$

dar cum

$$\hat{Y} = a + bX$$

rezultă, după înlocuire şi reducerea termenilor identici, că

$$\Delta y = b\,\sigma_x$$

Inlocuind acum pe b din formula (3), obţinem

$$\Delta y = r\,\frac{\sigma_y}{\sigma_x}\,\sigma_x$$

sau, simplificînd:

$$\Delta y = r \cdot \sigma_y$$

Deci variaţia lui $\hat{Y}$ este egală cu r abateri standard σ_y.

Aşadar, coeficientul de corelaţie r arată cu cîte abateri standard proprii, σ_y, variază Y atunci cînd X se modifică cu o

abatere standard σ_x, acceptînd că între cele două variabile există o relaţie de dependenţă lineară dată de dreapta de regresie a lui Y asupra lui X.

Să revenim acum la mărimea E_y. Am văzut că ea reprezintă suma pătratelor distanţelor de la valorile Y empirice la cele calculate prin ecuaţia dreptei de regresie; deci, într-un fel, ea ne dă suma pătratelor erorilor ce se comit atunci cînd înlocuim dependenţa statistică a lui Y faţă de X prin dependenţa funcţională lineară. Dacă împărţim pe E_y la n, numărul indivizilor, şi scoatem radicalul (pentru că diferenţele sînt la pătrat), obţinem o mărime

$$S_y = \sqrt{E_y/n}$$

numită eroare medie pătratică. Semnificaţia ei este evidentă: ea ne va indica abaterea medie a valorii efective, y_i, a unui individ de la valoarea calculată. Se poate, în plus, arăta că:

$$S_y^2 = \sigma_y^2 (1 - r^2) \qquad (5)$$

Se observă imediat că dacă r este egal cu unitatea, atunci S_y este nul (deci punctele nu se abat deloc de la dreapta de regresie) şi reciproc. In situaţia în care r este egal cu zero, din (5) rezultă imediat că

$$S_y^2 = \sigma_y^2$$

deci eroarea este maximă, coincizînd cu varianta lui Y.

Din formula (5) mai rezultă o interpretare a coeficientului de corelaţie. Scoţînd:

$$1 - r^2 = \frac{S_y^2}{\sigma_y^2} \qquad (6)$$

înseamnă că mărimea subunitară $1-r^2$ ne arată ce fracţiune din σ_y^2 este reprezentată de S_y^2. Cum S_y^2 poate fi interpretat ca o măsură a dispersiei punctelor în jurul dreptei de regresie, deci ca o nedeterminare rămasă pentru variabila Y, în ciuda relaţiei suprinse

în ecuaţia de regresie, atunci $1-r^2$ reflectă <u>proporţia din nedeterminarea totală</u> (σ_y^2) <u>rămasă încă.</u> Sau, scoţînd pe r^2 din (6),

$$r^2 = \frac{\sigma_y^2 - S_y^2}{\sigma_y^2}$$

rezultă că r^2 arată cu ce fracţiune se reduce varianţa lui Y prin introducerea dreptei de regresie.

In termeni mai puţin riguroşi (dar foarte încetăţeniţi în limbajul statisticii sociale, fără ca, din păcate, toată lumea să le înţeleagă cu adevărat sensul) se spune că r^2 arată cît din varianţa lui Y este explicată de X iar $1-r^2$ cît din ea rămîne în contul altor factori nespecificaţi.

Dacă inversăm acum datele problemei şi presupunem că X este funcţie de Y, vom ajunge la o dreaptă de forma:

$$\hat{X} = a_1 + b_1 Y \qquad (7)$$

Adică, pentru fiecare y_i va rezulta un $\hat{x}_i$ (un X calculat) şi determinarea coeficienţilor $\underline{a_1}$ şi $\underline{b_1}$ se va face punînd condiţia ca expresia

$$E_x = \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$$

sau

$$E_x = \sum_{i=1}^{n} (x_i - a_1 - b_1 y_i)^2 \qquad (8)$$

să fie minimă.

Ecuaţia (7), cu coeficienţii determinaţi din această condiţie, ne dă dreapta de regresie a lui X asupra lui Y sau cu Y variabilă independentă. Cele două drepte (1) şi (7) sînt, în general, distincte - adică ecuaţia (7) nu se deduce din (1) prin explicitarea lui X -, ele reflectînd cele două puncte de vedere după care se calculează distanţele ale căror sumă de pătrate se minimizează.

Este evident că toate considerațiile făcute în primul caz, precum și formulele date, rămîn valabile și aici, cu singura precizare că trebuie schimbat locul și rolul celor două variabile.

Regresia în cazul mai multor variabile. Să presupunem, de această dată, că avem definite m variabile $X_1, X_2, \ldots, X_m$, dintre care fie X_1 variabila dependentă. Cum fiecare individ ia acum cîte o valoare pentru fiecare variabilă, imaginile lor pot fi interpretate ca puncte într-un spațiu cu m dimensiuni. Ecuația de regresie se poate generaliza astfel:

$$\hat{X}_1 = a + b_2X_2 + b_3X_3 + \ldots + b_mX_m \qquad (9)$$

reprezentînd un hiperplan (generalizarea noțiunii de plan) în spațiul m-dimensional. Coeficienții b din ecuația (9) se numesc coeficienții regresiei parțiale, fiecare dintre ei arătînd cu cîte unități se modifică $\hat{X}_1$, dacă o variabilă X_j se modifică cu o unitate, celelalte fiind presupuse constante.

Pornind de la coeficienții b se pot defini așa-numiții coeficienți de corelație parțială între X_1 și oricare dintre variabilele din membrul drept. De pildă, dacă luăm corelația parțială între X_1 și X_2, coeficientul respectiv se va nota cu $r_{12.3\ldots m}$ și va avea semnificația unui coeficient de corelație între X_1 și X_2 dacă se presupune că $X_3, X_4, \ldots, X_m$ sînt menținute constante, arătînd deci legătura statistică între X_1 și X_2 "purificată" de intervenția celorlalte variabile explicitate.

Pentru calcularea coeficienților de corelație parțială nu este neapărat necesar să se determine ecuația de regresie. Există formule de calcul prin recurență, adică plecînd de la coeficienți de ordin inferior spre cei de ordin superior. Menționez că ordinul coeficienților de corelație parțială este dat de numărul variabilelor presupuse constante, deci, în notația coeficientului, de numărul de indici de după punct. Astfel coeficientul simplu între X_1 și X_2, pe care-l notăm acum r_{12} spre a pune în evidență cele două variabile, este un coeficient de ordinul zero. Un coeficient de ordinul unu se va obține "legînd" o variabilă, de pildă X_3 ; el se va nota $r_{12.3}$ și va arăta că e vorba de co-

relaţia dintre X_1 şi X_2 cînd X_3 este constantă. Dacă mai menţinem, în plus, constantă şi variabila X_4, atunci vom avea o corelaţie parţială de ordinul doi: $r_{12.34}$ etc.

Observaţie. In notaţia introdusă se desparte prin punct grupul indicilor variabilelor care se corelează de cel care precizează pe cele ce rămîn constante. Este clar că ordinea în care se scriu indicii din faţa punctului şi de după el nu are nici o importanţă; de aceea se preferă în scriere ordinea lor naturală. Adică $r_{12.34}$ este acelaşi lucru cu $r_{21.34}$ sau cu $r_{21.34}$ sau cu $r_{12.43}$ dar, evident, nu şi cu $r_{13.24}$ sau $r_{34.12}$ etc. fiindcă în aceste din urmă cazuri am schimbat poziţia indicilor în raport cu punctul.

In general, formula de recurenţă pomenită este următoarea:

$$r_{12.34\ldots m} = \frac{r_{12.34\ldots m-1} - r_{1m.34\ldots m-1}\, r_{2m.34\ldots m-1}}{\sqrt{(1 - r^2_{1m.34\ldots m-1})(1 - r^2_{2m.34\ldots m-1})}} \quad (10)$$

Cum spuneam, calculul efectiv porneşte de la coeficienţii de ordinul zero, adică cei obişnuiţi. De pildă:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r^2_{13})(1-r^2_{23})}}$$

Sau

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{\sqrt{(1-r^2_{14.3})(1 - r^2_{24.3})}}$$

etc.

Exemplu. Să presupunem că avem trei variabile între care se determină următoarele corelaţii:

$$r_{12} = 0{,}60; \quad r_{13} = 0{,}70\ ; \ r_{23} = 0{,}80$$

şi ne interesează relaţia dintre X_1 şi celelalte două variabile. Corelaţiile r_{12} şi r_{13} sugerează legături destul de strînse între prima variabilă şi celelalte două şi aproximativ de aceeaşi intensitate (o,6o şi o,7o). Făcînd constantă variabila X_3 şi aplicînd formula de recurenţă se obţine:

$$r_{12.3} = o{,}o9$$

Schimbînd apoi locul variabilelor X_2 şi X_3, se ajunge la:

$$r_{13.2} = o{,}46$$

Prin urmare, corelaţia dintre X_1 şi X_2 dispare aproape complet atunci cînd controlăm factorul X_3 ($r_{12.3}$ este foarte aproape de zero), pe cînd cea dintre X_1 şi X_3 rămîne încă la o valoare destul de ridicată - chiar dacă mai mică decît cea iniţială - cînd facem constant pe X_2. Din corelaţia iniţială foarte mare dintre cele două variabile "explicative" X_2 şi X_3, se putea deduce, de fapt, că relaţiile fiecăreia dintre aceste variabile cu X_1 vor fi puternic afectate de intervenţia celeilalte variabile. Iată că mărimea $r_{12.3}$ ne arată că aproape întreaga corelaţie iniţială de o,6o este datorată faptului că X_3 este puternic corelat cu X_2 şi X_1.

Un ultim indicator pe care doresc să-l mai prezint - deşi mai există şi alţii foarte importanţi - este coeficientul corelaţiei multiple. După notaţiile noastre din ecuaţia (9) coeficientul corelaţiei multiple, scris $R_{1(23...m)}$, reprezintă corelaţia globală dintre variabila X_1 şi celelalte variabile ai căror indici sînt menţionaţi în paranteza care apare la R. Indicatorul măsoară ca şi coeficientul simplu de corelaţie, gradul de precizie cu care ecuaţia de regresie (9) reproduce valorile lui X_1. Complementar, mărimea $1 - R^2$ reflectă partea din varianţa lui X_1 rămasă sau fracţiunea acestei varianţe "neexplicată" de $X_2, X_3, \ldots, X_m$.

Coeficientul corelaţiei multiple se poate scoate din formula de mai jos, care explicitează tocmai această mărime $1 - R^2$:

$$1 - R^2_{1(23...m)} = (1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})\ldots(1-r^2_{1m.23...m-1}) \qquad (11)$$

De pildă, revenind la cele trei variabile între care am dat mai sus coeficienţii de corelaţie simplă, vom aplica formula:

$$1 - R^2_{1(23)} = (1-r^2_{12})(1-r^2_{13.2})$$

Si atunci:

$$1 - R^2_1(23) = (1-0,36)(1-0,2116) = 0,5046$$

$$R^2_{1(23)} = 1 - 0,5046 = 0,4954$$

şi

$$R_{1(23)} = 0,704$$

Se observă, mai întîi, că $R_{1(23)}$ este doar cu foarte puţin mai mare decît r_{13}, ceea ce este în concordanţă cu rezultatul anterior conform căruia corelaţia parţială între X_1 şi X_2, cu X_3 constant, este foarte mică; adică introducerea lui X_2 în analiză nu aduce un aport sensibil în determinarea lui X_1 faţă de ceea ce ne dă X_3 singur. Pe de altă parte, se vede că, folosind expresia uzuală, din valoarea o,4954 a lui $R^2_{1(23)}$, variabilele X_2 şi X_3 "explică" _ şi nu voi înceta să pun în ghilimele acest cuvînt - 49,54% din varianţa lui X_1.

O proprietate normală a coeficientului corelaţiei multiple care poate s-a subînţeles din cele de mai sus, dar oricum trebuie subliniată, este aceea că $R_{1(23...m)}$ este mai mare decît oricare corelaţie simplă r_{1i} = 2,3,...m. Din comparaţia lui cu aceste din urmă valori se poate intui care este, cu aproximaţie, ordinul de mărime a corelaţiilor parţiale.

Cred că şi din sumarele consideraţii făcute în capitolul de faţă se poate înţelege ce rol important are analiza de corelaţie în studierea fenomenelor sociale cuantificabile. Calculul corelaţiilor simple ne oferă o primă imagine a intensităţii relaţiei dintre factorii analizaţi. Cum, de obicei, raporturile de determinare dintre fenomenele sociale sînt foarte complexe, deci este greu de găsit situaţii în care să se studieze o relaţie doar între două variabile, putînd face abstracţie de restul factorilor, este necesar

ca următorul pas al analizei să-l constituie calculul corelaţiilor parţiale şi multiplă. Analogia cu analiza multivariată a caracteristicilor calitative este evidentă; ambele proceduri au rolul de a suplini metoda experimentală efectivă, căci aceste analize de raporturi în care anumiţi factori se menţin constanţi nu fac decît să reproducă schema logică a experimentului.

Totuşi, atrag din nou atenţia că ne aflăm pe tărîmul statisticii, că indicatorii calculaţi aici nu sînt decît nişte instrumente prin care putem verifica (mai corect, infirma) anumite ipoteze teoretice privind determinaţiile dintre fenomene sau, cel mult, sugera unele legături între ele. De exemplu, oricît ar fi de bogat lotul de variabile auxiliare luat pentru a preciza relaţia dintre X_1 şi X_2, nu vom fi niciodată siguri că indicele de corelaţie parţială $r_{12.34...m}$ nu se modifică radical prin introducerea unei noi variabile, omisă pînă atunci. Rămîne sarcina teoriei sociologice de a analiza conţinutul variabilelor manipulate, de a stabili schemele explicative necesare.

PARTEA A IV-A

PROBLEMATICA EŞANTIONĂRII

Cap.XI. PROCEDEE DE EŞANTIONARE

1. Principiile cercetării selective

După cum am precizat în mai multe rînduri, în cercetările sociologice sîntem, de cele mai multe ori, în situaţia de a investiga proprietăţile, însuşirile etc. ale unui număr mare de unităţi (obiecte-suport, cum le-am numit), adică avem de a face cu o abordare cantitativă la nivelul obiectului, ceea ce duce la folosirea instrumentelor statistice. Este de subliniat că, în toate aceste cazuri, indiferent de natura obiectelor purtătoare de însuşiri, prima operaţie ce trebuie realizată, pentru a putea desfăşura în bune condiţiuni cercetarea, este delimitarea, circumscrierea cît mai exactă a populaţiei, respectiv a universului de obiecte asupra cărora cercetătorul îşi îndreaptă atenţia. Am precizat în capitolul referitor la etapele cercetării atît importanţa unei delimitări corecte a populaţiei, cît şi dificultăţile practice legate de această operaţie şi nu mai revin asupra lor. Ceea ce ne interesează în momentul de faţă este doar faptul că înaintea oricărei cercetări efective de teren trebuie ca cercetătorul să posede o delimitare a populaţiei ce va fi investigată şi asupra căreia să fie valabile concluziile ce rezultă în urma cercetării.

In situaţia în care toţi indivizii din populaţia astfel determinată sînt direct investigaţi, cercetarea se numeşte completă. Dacă însă populaţia este foarte mare, depăşind posibilităţile sociologului de a o imvestiga exhaustiv într-un timp util, acesta se va mulţumi să intre în contact direct doar cu o parte din universul de indivizi şi atunci cercetarea se numeşte selectivă. Partea din populaţie aleasă spre investigare va fi numită eşantion, iar totalitatea operaţiilor prin care se construieşte eşantionul

poartă numele de eşantionare sau selecţie.

Chiar în acest punct iniţial al expunerii simt nevoia să fac o precizare de maximă importanţă. A efectua o cercetare selectivă nu înseamnă a reduce populaţia sau universul cercetării; ea înseamnă numai că informaţia referitoare la populaţie IN TOTALITATEA SA nu se recoltează de la toţi indivizii, ci numai de la o parte dintre ei. Principiul gnoseologic care stă la baza cercetărilor selective este ideea că întotdeauna o parte reflectă, într-o anumită măsură, întregul. Principiul practic ce ghidează activitatea propriu-zisă de eşantionare îl constituie maximizarea măsurii în care partea reflectă întregul.

Nici ideea, nici practica eşantionării nu sînt lucruri noi. Ele există, s-ar putea spune, de cînd există activitate umană şi nu sînt, prin urmare, caracteristice doar cercetărilor sociologice. Eşantionări se efectuează atît în cadrul diferitelor cercetări de natură ştiinţifică, cît şi în activitatea curentă, de zi cu zi, a oamenilor. O gospodină care gustă supa spre a vedea cît e de sărată, un geolog care ia o mostră de roci dintr-un munte sau astronautul de pe Lună, asistenta medicală ce recoltează o cantitate mică de sînge, controlorii tehnici de calitate ce analizează amănunţit o parte din piesele produse într-o uzină etc. toţi aceştia efectuează, în fond, cercetări selective.

După cum se vede şi din exemplele date, uneori cercetarea selectivă este impusă de realitate, fiind imposibil un studiu complet (gospodina ar trebui să mănînce întreaga oală cu supă, astronautul să aducă Luna pe Pămînt etc.) Dar şi în cazurile în care este teoretic posibilă studierea completă a populaţiei, eşantionarea îşi dezvăluie o serie de avantaje incontestabile, care o fac foarte larg răspîndită în toate domeniile ce utilizează analiza statistică.

Sintetizînd avantajele aduse de cercetările selective în raport cu cele complete, cel puţin pentru domeniul sociologiei, se poate afirma că acestea apar pe următoarele paliere principale:

a) Avantaje de ordin practic-economic. Studiile selective se realizează, evident, la un cost financiar mai scăzut decît cele complete, fac apel la un număr mai mic de persoane calificate sau auxiliare de cercetare, realizîndu-se deci şi o economie de muncă.

b) Avantaje de ordin aplicativ. Rezultă, în principal, din scurtarea timpului necesar realizării cercetării, ceea ce-i conferă acesteia posibilitatea de a oferi soluţii prompte. Asemenea gen de avantaje este extrem de important mai ales în domeniul ştiinţelor sociale, unde schimbările se produc cu mare rapiditate şi nu o dată rezultatele cercetătirlor îşi pierd actualitatea pînă în momentul obţinerii sau publicării lor. Aşa de pildă, datele recensămintelor (exemplu clasic de cercetare socială completă) apar tipărite după 2-3 ani sau chiar mai mult de la data efectuării lor, cînd ele capătă doar, ca să nu zic valoare istorică, o semnificaţie de punct de referinţă pentru datele demosociale; dar este evident că utilitatea lor ar fi cu totul alta dacă cercetătorii şi organele decidente le-ar avea la dispoziţie mai rapid.

c) Avantaje de cunoaştere. Dacă cele spuse mai sus sînt chestiuni foarte simple şi evidente, avantajele de cunoaştere ale studiilor selective par mai neobişnuite, avînd în vedere că, în principiu cercetarea unui eşantion nu poate să dea decît o informaţie incompletă, mai săracă decît cea rezultată din studiile exhaustive. Si totuşi, nu rareori, în practică, lucrurile se inversează. Intr-adevăr, studierea unui lot mai restrîns de indivizi face posibilă utilizarea unui personal auxiliar (de exemplu, operatori de anchetă) mai puţin numeros, ce poate fi temeinic instruit, verificat (din punctul de vedere al aptitudinilor necesare cercetării, al trăsăturilor morale şi de personalitate etc.) şi controlat în activitatea concretă de teren. Prin urmare, se pot reduce substanţial erorile ce apar în procesul de culegere a datelor, erori care, în fapt, ocupă, de regulă, locul cel mai important în rîndul factorilor "perturbatori" ai oricărei cercetări sociologice.

Pe de altă parte, numărul mai mic de indivizi investigaţi dă posibilitatea, ca în aceleaşi condiţii de timp şi de cheltuieli, să se realizeze o cercetare de mai mare adîncime. Si timpul dar şi personalul restrîns şi bine instruit oferă şanse de a se aborda aspecte mai complexe, de a se aplica tehnici mai rafinate şi deci de a se ajunge la rezultate mai consistente, de esenţă, atunci cînd investigăm doar un eşantion. In plus, prin posibilitatea de a aplica

tehnici diferite aceluiaşi lot de indivizi, rezultatele obţinute pot fi controlate, coroborate şi astfel se pot elimina încă o serie de erori. Una este, de exemplu, informaţia oferită de o anchetă ce foloseşte un chestionar cu 20-25 de întrebări foarte simple, cum se întîmplă cînd populaţia e mare şi se face un studiu exhaustiv al ei, şi alta cea dată de un chestionar bogat, aplicat pe îndelete, capabil să surprindă mult mai multe aspecte, din corelarea cărora se pot degaja concluzii de mai mare fineţe.

In sfîrşit, tot pe linia unor asemenea avantaje se află şi cele legate de prelucrarea datelor în timp optim. Este limpede că, într-un acelaşi interval de timp, prelucrarea unui număr imens de chestionare, să zicem, nu poate să includă decît operaţii simple de tabulare, pe cînd dacă volumul e mai redus şi prelucrarea poate fi mai rafinată, timpul şi fondurile materiale la dispoziţie dînd posibilitatea unor operaţii şi analize statistice suplimentare. Desigur că însăşi pregătirea pentru prelucrare (verificarea chestionarelor, post_codificarea etc.) se poate face în condiţii mult mai bune cînd numărul acestora este mai redus.

Pentru realizarea operaţiilor practice de eşantionare, se ridică trei probleme de bază legate de:

a) mărimea eşantionului

b) reprezentativitatea acestuia şi

c) modul de alegere a elementelor constitutive.

Toate aceste trei aspecte sînt strîns intercorelate şi este dificilă o analiză separată a lor. Voi începe cu o tratare sumară a primelor două,- asupra cărora voi reveni cu determinări mai precise în capitolul următor -, urmînd apoi să descriu mai pe larg pe cel de al treilea, în concordanţă deci cu titlul acordat capitolului de faţă.

Să luăm ca punct de plecare mărimea eşantionului şi să vedem care sînt factorii ce concură la alegerea unui eşantion cu o dimensiune dată. Cei mai importanţi par a fi următorii patru:

(1) Caracteristicile populaţiei Aşa cum am mai spus, asupra unei populaţii se studiază seturi de însuşiri, caracteristici etc. In funcţie de acestea populaţia noastră poate manifesta un grad mai mare sau mai mic de omogenitate. Pentru studierea caracteristicilor faţă de care populaţia este omogenă, va fi suficient un eşantion de mărime mică. Ca să reiau un exemplu an-

terior, gospodina nu are nevoie decît de o linguriţă de supă spre a-şi da seama dacă e sau nu suficient de sărată (însuşirea ce se analizează), dat fiind că întreg conţinutul calei este omogen şi bine "amestecat" şi nu este de aşteptat ca într-o parte supa să fie mai sărată şi în alta mai puţin. Grupurile sociale evident că nu se caracterizează printr-o asemenea omogenitate, chiar în cazul unor variabile foarte simple. Si nu numai că eterogenitatea este, în general, mare, dar populaţia nu este nici atît de bine amestecată ca supra în oală, căci, aşa cum se ştie, există o clară tendinţă ca tipurile asemănătoare să se grupeze (Moser). Acest lucru ridică mult riscul de a trage concluzii pe baza unui eşantion restrîns, care n-ar fi capabil să acopere întreaga arie a unor astfel de grupări.

Cum în cercetările efective de teren se urmăresc întotdeauna mai multe caracteristici, va trebui să se ia în considerare pentru a stabili mărimea eşantionului aceea care are gradul de dispersie maxim. In cazul caracteristicilor cantitative, aşa cum s-a văzut în capitolele anterioare, dispersia poate fi măsurată prin indicatori numerici - în speţă, abaterea standard - indicatori care vor fi legaţi direct proporţional de mărimea eşantionului.

(ii) <u>Gradul de detaliere a analizei</u>. Este vorba aici de un lucru simplu şi evident, dar care nu este luat întotdeauna în considerare. Un volum al eşantionului poate fi suficient de mare pentru descrierea unei caracteristici în ansamblu, dar nu şi pentru realizarea unor comparaţii între situaţiile unor subeşantioare, delimitate cu o altă caracteristică. De exemplu, vom avea nevoie de un eşantion pentru a studia, să spunem, numărul mediu de camere ce revin unei familii din municipiul Cluj-Napoca şi de un altul, mult mai mare, dacă urmărim să estimăm aceeaşi mărime dar la nivelul cartierelor oraşului.

Aici se cuvine făcută o precizare importantă a cărei semnificaţie - ca de altfel şi a altor afirmaţii de mai înainte - se va dezvălui treptat pe parcursul acestui capitol şi al următorului. Nu există o reţetă universală pentru a spune cît de mare trebuie să fie un eşantion şi, mai ales, nu există o reţetă care să ne dea mărimea eşantionului <u>ca procent din numărul indivizilor ce compun populaţia.</u>

Mai întîi că - şi cei cu o oarecare perspicacitate, dar fără cunoştinţe statistice, vor sesiza-o cu surprindere - printre factorii principali ce determină mărimea eşantionului nu intervine mărimea populaţiei. Acest lucru poate părea paradoxal, dar el primeşte o demonstraţie riguros matematică. (De fapt, mărimea populaţiei influenţează pe cea a eşantionului doar într-un caz mai rar întîlnit în sociologie, cînd cele două dimensiuni sînt apropiate. Dacă se întîlnesc totuşi asemenea cazuri, e recomandabil să se studieze populaţia exhaustiv, căci întreg arsenalul teoretico-matematic al eşantionării porneşte de la principiul că se lucrează cu populaţii mai.) In al doilea rînd, că, din punct de vedere al problematicii eşantionării, egalitatea unor mărimi relative nu este totdeauna operantă. Mai exact, un eşantion de 5%, de exemplu, dintr-o populaţie de 10.000 unităţi nu are nici pe departe aceeaşi putere de reflectare a întregului ca un eşantion tot de 5% dar dintr-o populaţie de 1.000.000 de unităţi.

Am făcut aici aceste precizări, legate chiar de exemplul evocat mai sus, pentru că efectiv mi s-a întîmplat ca un beneficiar să nu poată înţelege că un eşantion bun pentru municipiul Cluj, în ansamblu, nu mai este bun pentru un cartier oarecare, deşi din întreg oraşul şi din fiecare cartier al său a intrat în eşantion exact acelaşi procent din familii.

(iii) Reprezentativitatea eşantionului . Este probabil noţiunea fundamentală în cercetătirle selective şi de aceea ar fi fost poate mai normal să încep cu ea sau chiar ca întreg setul acestor consideraţii să se raporteze la ea şi nu la dimensiunea eşantionului. Am ales totuşi această din urmă noţiune ca "punct de sprijin" pentru motivul că e mai uşor de înţeles şi pentru că reprezentativitatea va primi o definiţie riguroasă, sub formă matematică, doar în capitolul ce urmează. Pentru moment, vom spune că reprezentativitatea unui eşantion este gradul de concordanţă între caracteristicile definite pe acesta (ca parte) şi cele similare definite la nivelul întregii populaţii (ca întreg). Deci reprezentativitatea este, altfel spus, măsura în care partea reflectă întregul.

Chiar şi din această definiţie imprecisă rezultă un

lucru fundamental, pe care puţină lume îl cunoaşte şi îl înţelege cu adevărat: nu există eşantionare reprezentative şi nereprezentative, ci există doar eşantioane mai mult sau mai puţin reprezentative pentru o populaţie şi o caracteristică a acesteia date. Orice eşantion posedă o reprezentativitate diferită de zero şi nici un eşantion nu este 100% reprezentativ sau, mai corect spus, probabillitatea de a se întîmpla acest din urmă lucru este practic nulă.

Intre mărimea eşantionului şi reprezentativitatea sa există desigur o dependenţă directă, în sensul că, toţi ceilalţi factori fiind egali, pe măsură ce creşte volumul eşantionului sporeşte şi gradul său de reprezentativitate, reciproc, pentru a spori reprezentativitatea, în condiţii egale, trebuie mărit volumul eşantionului. Un eşantion de mărime nulă - dacă se poate vorbi de aşa ceva - va avea o reprezentativitate nulă şi unul egal cu întreaga populaţie va avea una completă (să-i zicem 100% Dar această dependenţă este departe de a fi lineară, deci creşterile pe una din cele două dimensiuni menţionate nu antrenează creşteri proporţionale pe cealaltă. Spre a sugera cît de cît forma dependenţei dintre cele două mărimi, am construit graficul din figura 1, în care se vede că la o creştere a valorilor mici ale mărimii eşantionului corespund creşteri rapide ale reprezentativităţii; dimpotrivă, după ce numărul de indivizi depăşeşte o anumită zonă, marcată de segmentul AB, practic orice spor al acestuia aduce o creştere infirmă a reprezentativităţii. Prin urmare, cîştigurile realizate de pe urma mărimii exagerate a eşantionului sînt minore şi ele, de regulă, sînt contrabalansate de creşterea gradului de eroare implicat de o mulţime de factori atunci cînd cîmpul de investigaţie se lărgeşte (şi asta fără a mai vorbi de ridicarea costului cercetării şi de alte impedimente practice.).

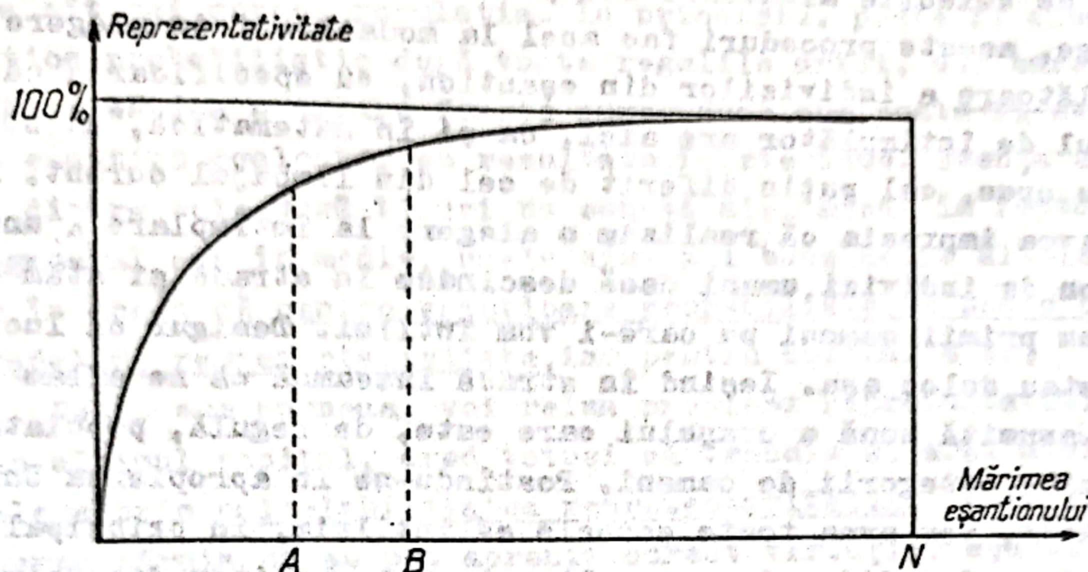


Fig. 1 Relaţia dintre volumul eşantionului şi reprezentativitatea sa (N = volumul populaţiei)

(iv) Procedeul de eşantionare. Dat fiind că nivelul de reprezentativitate al eşantionului (atunci cînd el poate fi calculat) este în funcţie de tehnica de selecţie, este clar că şi mărimea eşantionului depinde de acest factor. La un acelaşi nivel de reprezentativitate dat, folosind procedee de eşantionare diferite, vom avea nevoie de eşantioane de mărimi diferite.

2. Tipuri de eşantioane

Una dintre problemele de bază ale cercetării selective este aceea a procedeelor practice prin care se aleg din populaţie elementele constituive ale elantionului ce urmează investigat. In principiu, toate lucrările metodologice admit existenţa a două mari clase (tipuri sau genuri) de eşantioane: cele alese printr-o procedură probabilistă, numite eşantioane aleatoare şi cele care nu fac apel la astfel de proceduri, eşantioane nealea-

toare.

Primul lucru asupra căruia ne vom opri este sensul noţiunii de selecţie aleatoare sau probabilistă. După cum se poate înţelege, aceste proceduri fac apel la modalităţi de alegere întîmplătoare a indivizilor din eşantion, cu specificarea că termenul de întîmplător are aici, ca şi în matematică, un sens, dacă nu opus, cel puţin diferit de cel din limbajul curent. Am putea avea impresia că realizăm o alegere la întîmplare a unui eşantion de indivizi umani dacă descindem în stradă şi stăm de vorbă cu primii oameni pe care-i vom întîlni. Desigur că lucrurile nu stau deloc aşa. Ieşind în stradă înseamnă că ne aflăm într-o anumită zonă a oraşului care este, de regulă, populată cu anumite categorii de oameni. Postîndu-ne în apropierea Universităţii, vom avea toate şansele să întîlnim, în principal, studenţi, cadre didacrice sau alt personal muncitor din Universitate. Similar se petrec lucrurile dacă locul nostru de descindere este o zonă industrială, comercială etc. Apoi, la o anumită ori din zi, şansele sînt foarte mari să întîlnim sau să nu întîlnim pe stradă anumite categorii de oameni. Nu mai vorbesc de faptul că fiecare dintre noi, oricîtă obiectivitate şi-ar impune are tendinţa să aleagă şi să respingă anumite tipuri de persoane. Aşadar, este limpede că o asemenea alegere "întîmplătoare" nu poate satisface cerinţele unei cercetări selective seriaose, deoarece eşantionul este departe de a surprinde structurile de bază ale populaţiei şi deci este de aşteptat că şi în alte privinţe, referitor la alte caracteristici, chiar dacă e vorba de opinii, motivaţii etc., el nu va da o imagine suficient de corectă a realităţii.

Intîmplarea matematică trebuie căutată prin procedee riguroase, care să elimine orice alegere subiectivă sau orice intervenţie a unui factor perturbator cu acţiune constantă (cum este abordarea oamenilor pe stradă). In general, se apreciază că pentru realizarea unui eşantion probabilist este nevoie ca fiecare element din populaţie să aibă o şansă CALCULABILĂ şi NONNULĂ de a fi ales. Orice eşantionare care nu respectă aceste cerinţe va fi considerată nealeatoare.

Diferenţierea conceptuală a celor două tipuri de selecţie este foarte importantă, chiar dacă, din punctul de vedere al

eficienţei practice, este imposibil să se indice vreo superioritate absolută a unui gen sau a altuia. Vorbind de superioritate, mă refer desigur la posibilitatea pe care o are eşantionul de a reflecta cît mai corect populaţia. In principiu, poate fi ales un eşantion probabilistic după toate regulile artei, dar care, întîmplător să dea o eroare foarte mare, după cum poate fi studiat un eşantion nealeator cu rezultate foarte bune. Esenţa deosebirii dintre cele două tipuri nu constă nici măcar în faptul că, în general sau în medie, unele sînt mai bune decît altele, ci doar în aceea că pentru eşantioane probabiliste se poate calcula gradul de reprezentativitate, iar pentru celelalte nu.

Deşi, cum spuneam, voi relua problema reprezentativităţii în ultimul capitol, cred totuşi că trebuie să arăt aici ce se înţelege prin calculabilitatea reprezentavităţii, lucru fără înţelegerea căruia nu se pot apresia corect virtuţile eşantionării alteatoare. Să presupunem că prin cercetarea selectivă urmărim să estimăm o valoare v a unei caracteristici din populaţie (de exemplu, vîrsta medie, retribuţia medie, procentul celor ce exprimă o opinie etc.). Cercetînd efectiv indivizii din eşantion, vom da peste o valoare corespunzătoare v', care, în general, este diferită de v. Este clar că un eşantion este cu atît mai reprezentativ cu cît v' se apropie mai mult de v. S-ar părea că acea calculabilitate de care pomeneam n-ar consta în altceva decît în efectuarea diferenţei dintre v şi v' (abaterea lui v' de la v). Numai că v este o mărime necunoscută şi prin urmare este imposibil să se calculeze această diferenţă. Cea ce oferă eşantionarea aleatoare este posibilitatea de a estima această diferenţă în termeni de probabilităţi. Mai exact, pentru orice eşantion aleator, se poate afirma cu o anumită probabilitate P, calculabilă, că mărimea din eşantion, v', nu diferă de cea adevărată (din populaţie), v, cu mai mult decît o mărime Δ, ce poate fi făcută oricît de mică, aşa cum probabilitatea P poate fi făcută oricît de mare.

Să trecem în continuare la prezentarea pe scurt a principalelor tipuri de eşantioane, începînd cu cele probabiliste. Spun pe scurt deoarece problematica acestui paragraf este extrem de vastă şi nu-i voi putea face decît un succint rezumat. Pentru mai multe detalii, fac trimitere la lucrarea lui C.A.Moser (1967)

tradusă în româneşte, unde se pot găsi informaţii şi precizări suplimentare, dezvăluind, în acelaşi timp, şi rolul important pe care această carte l-a avut în organizarea ideilor din capitolul de faţă.

2.1. Eşantionarea aleatoare

Tehnicile de selecţie aleatoare, sînt la rîndul lor, clasificate în mai multe tipuri principale ce se aleg în funcţie de necesităţile studiului, de răspîndirea – respectiv, gruparea – populaţiei în teritoriu, de baza informaţională existentă înainte de începerea cercetării etc. Si dacă am pomenit de baza informaţională, să mai fac o precizare importantă: <u>o procedură de eşantionare aleatoare – în sensul cel mai riguros al termenului – nu poate fi utilizată decît atunci cînd există un cadru de eşantionare corect întocmit</u>. Prin cadru de eşantionare se înţelege o listă completă a populaţiei din care să se extragă eşantionul sau o altă informaţie despre populaţie echivalentă cu aceasta (de exemplu, o hartă pe care să fie marcate toate unităţile). Se înţelege că întocmirea unui cadru de eşantionare presupune, pe lîngă o definiţie operaţională riguroasă a populaţiei, şi existenţa unor posibilităţi practice pentru cercetător de a identifica efectiv (şi nu potenţial sau în principiu) fiecare element al populaţiei. Lucrul acesta nu este deloc simplu şi dacă sociologul nu posedă dinainte un asemenea cadru de eşantionare, din stocul de date existente la anumite organe sau instituţii (consilii populare, miliţie, poştă etc.) este foarte puţin probabil ca el însuşi să aibă posibilitatea de a şi-l construi.

Procedeele de bază utilizate în eşantionarea aleatoare sînt:

A. Procedee elementare (eşantionarea simplă aleatoare)

B. Eşantionarea prin stratificare

C. Eşantionarea multistadială

D. Eşantionarea multifazică

Să trecem la prezentarea lor în această ordine.

<u>A</u>. <u>Procedee elementare</u>. Cel mai simplu, ca principiu, procedeu elementar este aşa-numita <u>metodă a loteriei</u>: indivizii populaţiei care vor constitui eşantionul sînt "traşi la sorţi" cu

ajutorul unei urne, unde fiecare unitate este prezentă, de exemplu sub forma unui bilețel sau a unei bile. Toată lumea cunoaște precauțiunile ce se iau la astfel de extrageri (cum ar fi cele de la loto sau pronoexpres): bilele trebuie să aibă dimensiuni și mase identice, urna trebuie să permită o cît mai bună amestecare a lor etc. Toate aceste rigori sînt impuse de obiectivul fundamental - care rămîne valabil și în eșantionarea în scopuri științifice -, și anume acela de a asigura tuturor unităților <u>șanse egale</u> de a fi extrase.

Dacă populația este formată din <u>N</u> indivizi, atunci este evident că, în condiții ideale, probabilitatea ca un individ să fie extras este 1/N. Se observă, că după extragerea unui individ, probabilitățile pentru ceilalți se schimbă puțin: înainte de extragerea celui de al doilea în urmă se află N-1 indivizi și deci probabilitatea devine 1/N-1; înainte de extragerea celui de al treilea, probabilitatea va fi 1/N-2 ș.a.m.d. Deci șansele de a fi prezenți în eșantion, pentru indivizii rămași în urnă, c cresc pe măsură ce se înaintează cu aceste extrageri. Pentru a se menține mereu aceeași probabilitate, se poate proceda la reintroducerea în urnă a fiecărui individ extras, totalul fiind atunci mereu N. In felul acesta se face distincția între două tipuri de selecții: <u>repetată</u> (cînd se reintroduc în urnă indivizii extrași) și <u>nerepetată</u>. Diferența nu are o importanță practică deosebită decît atunci cînd volumul populației este relativ mic, adică atunci cînd scăderea unui număr de indivizi din total (număr ce merge pînă la volumul eșantionului minus unu) influențează sensibil probabilitățile 1/N-i. Teoretic, diferența între cele două proceduri își pune amprenta asupra formulelor de calcul ce intervin.

Utilizarea efectivă a metodei loteriei este foarte restrînsă, în cazul cercetărilor sociologice de teren. Motivul fundamental rezidă în numărul mare de indivizi ce formează, de regulă, populația și pentru care este imposibil să se asigure condițiile tehnice necesare unei "trageri la sorți" riguroase. A scrie mii, zeci sau sute de mii de bilețele perfect egale, a le amesteca suficient de bine pentru a elimina orice efect al factorilor ce au determinat ordinea introducerii lor în urnă, a le extrage apoi din urnă cu totul întîmplător sînt cerințe foar-

te greu de îndeplinit. Din acest motiv se caută substituirea acestor procedee cu altele, echivalente sau aproape.

Unul dintre acestea, de o mare rigurozitate şi, în acelaşi timp, cu o eficienţă practică deosebită constă în utilizarea tabelelor cu numere aleatoare. Aşa cum o arată şi numele e vorba de tabele ce conţin o mulţime de numere determinate de o manieră strict probabilistică. Trec peste detaliile de construcţie a lor, dar precizez că se utilizează metode foarte sofisticate şi nu este pur şi simplu o înşirurire de numere aşa cum îi vin cuiva în minte. Numerele pot fi formate din 4,5,6 sau mai multe cifre şi sînt aşezate, de regulă, sub forma unor coloane în felul următor:

452402	390946	...
419415	886954	...
963827	250162	...
710612	748522	...
981450	054556	...
...	...	...

Modul de utilizare a tabelelor este foarte simplu. După cum am spus, pentru orice eşantionare aleatoare posedăm sau trebuie să construim o listă a tuturor unităţilor din populaţie, listă pe care indivizii pot fi numerotaţi - şi de regulă, numerotarea se face chiar odată cu întocmirea listei - de la 1 la N. Alegerea unui eşantion de volum n revine la următoarele: se porneşte dintr-o zonă oarecare a tabelului de numere aleatoare, pe coloane sau pe linii, şi numerele întîlnite ne vor indica poziţia indivizilor din lista noastră ce vor intra în eşantion. Dacă tabelul conţine numere cu mai multe cifre decît are N, atunci se renunţă la primele cifre ale numerelor din tabel. Dar mai bine să recurgem la un exemplu concret. Să presupunem că populaţia noastră este formată din 7200 indivizi, din care vrem să selecţionăm 600. Folosind un tabel similar celui din care am extras numerele de mai sus, se vede că va trebui să renunţăm la primele două cifre, iar dintre numerele cu patru cifre rămase, de asemenea trebuie eliminate cele mai mari decît 7200. Pornind chiar de la primul număr scris pe coloane, primul individ extras va fi cel cu numărul 2402, al doilea cel cu numărul 3827 (s-a sărit peste

9415 care depăşeşte valoarea lui N), al patrulea 612 (zeroul din faţă nu se ia în considerare), al cincilea 1450 ş.a.m.d. pînă se completează suma de 600.

O formă de eşantionare - deseori folosită şi considerată a poseda atributul de aleatoare - este aşa-numita metodă a pasului. Reluînd exemplul anterior, cu presupunerea că trebuie să alegem 600 indivizi din 7200, rezultă, din împărţirea:

$$7200 : 600 = 12$$

că tot unul din 12 indivizi va fi ales în eşantion. Se ia atunci lista celor 7200 de unităţi, se alege la întîmplare un individ din primii 12 aflaţi pe listă (de pildă prin tragera la sorţi) după care numerele de ordine ale celorlalţi se vor obţine adăugind succesiv cifra 12 la numărul anterior determinat. Astfel, dacă am ales în eşantion pe al 5-lea de pe listă, următorul va purta numărul de ordine 17(5+12) etc. La sfîrşit vom obţine 600 de indivizim exact cît ne-am propus.

Valoarea

$$k = \frac{N}{n}$$

se numeşte pasul de eşantionare. Raportul poate duce la o valoare întreagă, aşa cum s-a întîmplat în cazul ilustrat, şi astfel procedura ne dă efectiv un eşantion de mărimea urmărită. Dacă pasul k iese o mărime fracţionară, atunci se poate lucra cu întregul imediat mai mic decît k - partea întreagă a lui k - sau cu cel imediat superior - partea întreagă a lui k+1. In prima situaţie, volumul eşantionului va depăşi cifra prevăzută, ceea ce înseamnă că dintre indivizii aleşi o parte trebuie eliminaţi. Acest lucru se poate realiza folosind tot metoda pasului selectînd din eşantionul rezultat în prima fază numărul suplimentar de indivizi. In al doilea caz, volumul eşantionului va fi mai mic decît cel prevăzut şi numărul indivizilor necesari spre completare se poate determina printr-o nouă selecţie asupra întregii populaţii, cu un pas corespunzător.

Metoda pasului nu este stricto sensu una aleatoare, deoarece odată cu alegerea primului individ (să zicem al i-lea) se determină exact şi ceilalţi, avînd rangurile i+k, i+2k, i+3k

etc. Deci pentru restul probabilitatea de a fi aleşi este nulă, ceea ce contrazice principiul însuşi al eşantionării aleatoare. Dar, în practică, metoda dă rezultate bune, mai ales atunci cînd listarea se face după un criteriu (de pildă cel alfabetic) ce nu are nici o legătură cu probleme ce urmează a fi studiate. In caz contrar, se pot produce unele erori sistematice. De exemplu, dacă cele 7200 unităţi sînt localităţi pentru care se urmăreşte estimarea mărimii medii şi dacă ele sînt înşirate în ordinea crescătoare a volumului lor demografic, atunci cînd elementele alese în eşantion poartă numere de ordine mici - de pildă, 1,13,25 etc. - se va subestima mărimea cercetată, iar cînd poartă numere mari (12, 24, 36 etc.) se va supraestima.

B. Eşantionarea prin stratificare. Este o procedură care dă rezultate mai bune în privinţa reprezentativităţii decît cea simplă aleatoare. Mai exact, la acelaşi volum al eşantionului, nivelul de reprezentativitate este mai ridicat în cazul eşantionării prin stratificare, Ideea de bază constă aici în divizarea populaţiei în straturi, în funcţie de una sau mai multe caracteristici, pentru fiecare strat realizîndu-se separat o eşantionare aleatoare. In felul acesta, mărimile subeşantioanelor corespunzînd diferitelor straturi vor fi proporţionale cu ponderea acestor straturi în populaţia totală, avînd garanţia că fiecare strat este fidel reprezentat în eşantion. Se limitează aşadar jocul şansei , fără ca prin aceasta să se încalce principiile eşantionării aleatoare.

Drept factori de stratificare se aleg, de preferinţă, acele variabile care au o influenţă importantă asupra fenomenelor studiate. Dacă este vorba de un univers de indivizi umani, asemenea factori vor fi: sexul, vîrsta, ocupaţia, şcolaritatea, nivelul veniturilor etc. Un rol similar îl poate juca apartenenţa la unităţi administrativ-teritoriale, la diferite grupuri sau instituţii sociale.

Să dăm şi un exemplu, pentru a înţelege mai clar semnificaţia şi modalitatea practică de realizare a procedurii. Să presupunem că dintr-o populaţie de persoane, în mărime de 100.000 , ocupate în industria unui judeţ, dorim să investigăm direct un eşantion de 1.000 indivizi. Grupînd persoanele ocupate pe categorii socio-profesionale, vom ajunge, să zicem, la situa-

ţia de mai jos:

muncitori	-	85.ooo
maiştri	-	7.5oo
tehnicieni	-	3.5oo
funcţionari	-	2.5oo
ingineri, economişti	-	1.5oo
Total	N =	1oo.ooo

Drept urmare, eşantionul nostru de 1.ooo de persoane va trebui să cuprindă unu la sută din fiecare categorie, adică: 85o muncitori, 75 maiştri, 35 tehnicieni, 25 funcţionari şi 15 ingineri şi economişti, reflectînd exact ponderea fiecărei ocupaţii în populaţie.

De multe ori aceste stratificări au drept scop nu doar o reprezentare corectă a straturilor, ci realizarea unui studiu comparativ privind situaţia fiecăruia dintre straturi. In astfel de cazuri, procedura nu mai poate fi utilizată exact în forma simplă ilustrată mai sus, căci anumite straturi de dimensiune redusă în populaţie vor fi <u>numeric</u> slab reprezentate în eşantion. Revenim, de fapt, la o problemă atinsă şi în paragraful precedent, care, de data aceasta, constă în faptul că dacă cei 85o de muncitori pot constitui un eşantion cu o reprezentativitate suficientă în raport cu populaţia muncitorilor, pentru restul categoriilor eşantioanele nu mai posedă această proprietate, deşi, ca pondere ele sînt egale (1%). In cazuri ca acesta, nu se vor putea face comparaţii între straturi, iar pentru realizarea unui asemenea obiectiv vor trebui mărite simţitor subeşantioanele de mică dimensiune. Dar atunci nu va mai apare corect imaginea globală, ea fiind evident deformată de suprareprezentarea unor straturi.

Pentru a împăca ambele cerinţe, se poate realiza studiul cu un eşantion deformat, în care fiecare subpopulaţie să fie reprezentată de un număr de indivizi corespunzător unei reprezentativităţi dorite pe subpopulaţii, iar apoi, la constituirea imaginii de ansamblu, mărimile estimate pe straturi se

vor pondera în funcţie de mărimea reală a stratului. Astfel, dacă în exemplul de mai sus, estimînd media cîştigurilor lunare pentru cele 5 categorii - din studiul unor eşantioane suficient de reprezentative pentru fiecare - obţinem valorile $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, $\bar{x}_4$ şi $\bar{x}_5$, media estimată pe întreaga colectivitate se poate calcula în felul următor:

$$\bar{x} = 0,850\bar{x}_1 + 0,075\bar{x}_2 + 0,035\bar{x}_3 + 0,025\bar{x}_4 + 0,15\bar{x}_5$$

Mai există, desigur, şi o altă posibilitate de ajustare a datelor globale. După studierea subeşantioanelor pentru fiecare strat se pot alege eşantioane mai mici, proporţionale cu talia stratului, care să se reunească pentru a obţine un eşantion fidel ansamblului. Studiul acestui eşantion va da estimările pentru întreaga populaţie. Preferabilă pare a fi prima metodă căci aceasta din urmă poate introduce erori suplimentare la selecţia a doua .

In fine, o observaţie ce se impune în legătură cu eşantionarea prin stratificare. Este clar că avantajele sale ar trebui să conducă la o largă utilizare a procedurii, numai că în realitate lucrurile stau altfel. Condiţa fundamentală - cunoaşterea structurilor populaţiei înainte de studierea ei - este rareori îndeplinită; statisticile noastre curente dau puţine informaţii care să poate fi utilizate de cercetarea sociologică şi asta nu numai pentru construirea unor eşantioane de înaltă precizie.

C. <u>Esantionarea multistadială</u>. Se foloseşte probabil mai des decît cea precedentădar nu din raţiuni de ameliorare a estimărilor, ci din altele, de natură practică. Se ştie că populaţiile statistice nu sînt distribuite uniform în teritoriu sau în alt spaţiu de definire a lor. Ele au tendinţa de a se concentra sub formă de grupuri mai mult sau mai puţin delimitate. Dacă e vorba de indivizi umani, cum se întîmplă de multe ori în studiile noastre, aceştia aparţin unor grupuri familiale, rezidenţiale, de muncă sau învăţătură etc. La rîndul lor, aceste grupuri de gradul întîi, să le zicem, se pot diviza în grupuri secundare, care la rîndul lor ... etc.

Principiul eşantionării multistadiale constă într-o selecţie <u>indirectă</u> a indivizilor ce formează eşantionul, prin in-

termediul selecţiei grupurilor la care aceştia aparţin. In primul stadiu al eşantionării se alege un eşantion de grupuri de gradul I; apoi din fiecare grup aici ales, se selectează un eşantion de grupuri de gradul II etc. pînă cînd se ajunge la ultimul nivel reprezentat de indivizi. In acest stadiu, se poate include în eşantion întreg grupul imediat superior nivelului individual, ceea ce se întîmplă, de regulă, cînd grupurile sînt de dimensiune redusă.

Un exemplu simplu ne va lămuri şi de această dată mult mai bine asupra procedurii decît o expunere abstractă. Să zicem că avem nevoie de un eşantion de studenţi la nivel de ţară. Cîteva paliere de grupare a studenţilor pot fi următoarele:

I. Centre universitare
II. Institute de învăţămînt superior
III. Facultăţi
IV. Grupe de studenţi

Am lăsat, după cum se vede, la o parte secţiile şi anii de studiu, care ar fi putut mări încă numărul nivelelor de grupare.

Eşantionarea multistadială constă, în cazul de faţă, în alegerea unui eşantion aleator de centre universitare; în cadrul centrelor alese se vor selecţiona institute de învăţămînt superior, din care se aleg facultăţi, apoi în acestea grupe şi, în fine, în cadrul grupei un lot de studenţi. Cum spuneam, neputem opri la penultimul stadiu, adică să includem în eşantion grupa de studenţi în întregime. Acest lucru este în deplină concordanţă cu raţiunea principală ce ghidează folosirea unei asemenea selecţii: efectuarea studiului asupra unei populaţii cît mai grupate, în scopul economisirii timpului şi a mijloacelor materiale. Este clar, în cazul de mai sus, că anchetatorii nu vor mai trebui să se deplaseze în toate centrele universitare, ci doar în cîteva, apoi şi aici vor merge doar în unele institute şi facultăţi. Si în alte cazuri de anchete se pot evidenţia avantajele practice ale acestei eşantionări. De pildă, dacă trebuie ales un eşantion dintr-un oraş mare, pe schiţa localităţii se pot trasa carouri de o anumită dimensiune, dintre care cîteva se aleg în mod aleator şi doar în interiorul acestora se va selecta eşantionul de indivizi.

Operaţiile practice legate de eşantionarea multistadială sînt ceva mai complicate decît au lăsat să se înţeleagă rîndurile de mai sus. In primul rînd se pune problema inegalităţii mărimii grupurilor; ea se rezolvă dînd şanse diferite fiecărui grup de a intra în eşantion, şanse proporţionale cu volumul său. Apoi, trebuie efectuate calcule ceva mai complexe legate de reprezentativitatea eşantionului. Aici trebuie reţinut că, la mărimi egale, eşantionul multistadial este mai puţin reprezentativ decît cel simplu aleator. Motivul este evident: alegînd grupuri, care sînt în număr mult mai mic decît cel al indivizilor, cresc şansele de a se omite unele categorii din populaţie.

D. Eşantionarea multifazică. Mărimea eşantionului depinde, printre altele, aşa cum spuneam, de caracteristicile populaţiei studiate şi de gradul de reprezentativitate urmărit. Dar reprezentativitatea trebuie, la urma urmei, calculată pentru fiecare caracteristică în parte, ea fiind invers proporţională cu dispersia variabilei în cauză. De aici rezultă că dacă se alege eşantionul de aşa manieră încît el să aibă o reprezentativitate satisfăcătoare în raport cu variabila cea mai dispersată, atunci pentru celelalte caracteristici mărimea eşantionului devine excesivă, plusul de precizie realizat neavînd o valoare deosebită pentru investigaţie. Pe de altă parte, în cadrul oricărei cercetări există probleme mai importante - sau cărora sociologul le ataşează o importanţă mai mare - şi altele mai puţin importante. Pentru primele se va urmări, de regulă, obţinerea unor estimări foarte precise, în timp ce ultimele sînt abordate satisfăcător şi dacă eşantionul este mai mic şi nu se realizează o evaloare foarte riguroasă.

In al treilea rînd, tot plecînd de la distincţia anterioară între aspectele cercetate pe criteriul importanţei, în cazul anchetelor, mai ales, se urmăreşte nu o dată ca unele chestiuni să beneficieze de un studiu mai intens; dacă, de exemplu, metoda de bază de culegere a datelor este ancheta exată pe un chestionar cu cvasitotalitatea întrebărilor închise şi cu formulări standardizate, se simte nevoia ca aspectele de bază să fie abordate şi prin alte metode, cum ar fi interviul, care să aducă un plus de cunoaştere pe tărîmul motivaţiilor, al atitudinilor etc.

Dar o asemenea investigaţie intensivă nu este posibilă pentru întreaga masă de indivizi din eşantion căci nu se dispune nici de timp, nici de personal calificat pentru interviuri, nici de posibilităţi de analiză şi prelucrare a unei asemenea cantităţi de informaţie nestandardizată.

Toate problemele semnalate mai sus duc la adoptarea unei tactici de cercetare specifice, surprinsă în titlul de eşantionare multifazică. E vorba, în linii mari, de faptul că se alege iniţial un eşantion de dimensiune mare, în cadrul căruia se realizează diferite faze ale cercetării; apoi din acesta se selecţionează un subeşantion pentru realizarea altor faze, eşantion ce, la rîndul lui, poate fi micşorat printr-o selecţie aleatoare pentru alte faze etc. De exemplu pe eşantionul mare putem aplica un chestionar simplu, cu puţine întrebări şi răspunsuri standardizate, în scopul de a decela unele structuri de bază ale populaţiei. In faza a doua, pe un subeşantion, putem aplica un chestionar mai bogat, cu întrebări deschise, de opinie şi atitudine; într-o fază ulterioară, pe un eşantion de mărime şi mai redusă se poate aprofunda în continuare cercetarea prin interviuri, bugete de timp, bugete de familie, observaţie etc.

Aceste faze, logic delimitate aici, se pot realiza simultan în timp. Adică , dacă e vorba de anchetă, operatorul primeşte instrucţiuni exacte la care dintre indivizii anchetaţi să aplice numai instrumentele prevăzute pentru prima şi a doua ş.a.m.d. Instrumentele, la rîndul lor, pot fi separat întocmite sau pe unul şi acelaşi formular vor fi marcate întrebările specifice diferitelor faze ale cercetării.

Desigur că, din punctul de vedere al preciziei eşantionării în fazele ulterioare, este preferabil ca lucrurile să se deruleze succesiv în timp şi nu simultan. Astfel, după culegerea datelor în prima fază şi după o prelucrare sumară a lor, vom avea posibilitatea să obţinem, pentru următoarele faze, eşantioane prin stratificare, folosind informaţia mai înainte dobîndită. In felul acesta, reprezentativitatea eşantioanelor mai mici din fazele ce urmează va creşte sensibil. O asemenea strategie a cercetării îi măreşte evident costul şi timpul de realizare şi deci considerentele practice pot preleva asupra celor de

natură cognitivă.

O precizare importantă se impune şi aici. Avînd de a face cu o eşantionare probabilistică, în situaţia cînd toate fazele sau mai multe faze se realizează simultan, este nevoie ca anchetatorii să primească instrucţiuni foarte exacte legate de identitatea persoanelor ce formează eşantioanele din fazele a doua, a treia etc. Adică, nu trebuie lăsată la latitudinea operatorilor alegerea subiecţilor chestionaţi în aceste faze, deoarece astfel se iese din cadrul procedurilor de selecţie probabilistă.

In încheierea acestui subparagraf, doresc să subliniez ideea, ce s-a putut deja întrezări pe parcursul său, că cele patru tipuri de eşantionare probabilistă nu se exclud unele pe altele şi deci că, în cadrul uneia şi aceleiaşi cercetări, putem avea o împletire a lor. Practic, orice combinaţie este posibilă: eşantion stratificat bistadial, eşantion stratificat bifazic etc. In astfel de cazuri, una dintre problemele principale este cea a calculului reprezentativităţii, fiecare procedură făcînd apel la alte formule. Spun aceasta pentru că nu o dată se folosesc astfel de procedee combinate de eşantionare, în timp ce calculele de reprezentativitate sînt făcute ca şi cînd eşantioanele ar fi aleatoare elementare.

2.2. Eşantionarea nealeatoare

Am arătat mai înainte că principiile de bază ale eşantionării aleatoare sînt: a) posibilitatea calculării şanselor fiecărui individ de a intra în eşantion şi b) faptul că aceste şanse nu sînt egale cu zero pentru nici o unitate a populaţiei. Ori de cîte ori o procedură de eşantionare încalcă unul din aceste două - sau ambele - principii ea va trebui considerată nealeatoare. Practic, asemenea proceduri pot căpăta forme de realizare concretă nelimitate şi este foarte greu să se facă o tipologizare a lor. Vom avea eşantioane neprobabilitste atunci cînd ieşim în stradă să chestionăm un număr de subiecţi, cînd obţinem răspunsuri la un chestionar publicat într-o revistă sau un ziar, dacă anchetăm spectatorii prezenţi într-un cinematograf sau teatru, dacă ne ducem într-o şcoală spre a aplica un chestionar

elevilor, cînd, dintr-un județ, alegem 2-3 localități rurale "tipice" pentru investigarea anumitor probleme, dacă ... și exemplele ar putea continua aproape la nesfîrșit, atît de multe variante este posibil să imaginăm sau să le găsim efectiv puse în practică.

Grija principală a oricărui cercetător serios care utilizează în investigație un eșantion nealeator va fi aceea de a limita pe cît posibil subiectivitatea în alegerea unităților ce formează eșantionul. Ingrădirea subiectivității de alegere se poate face prin stabilirea inițială a unor caracteristici pe care să le posede fievare individ ales. S-a ajuns astfel să se prefigureze și în cadrul selecției nealeatoare o procedură, ale cărei rezultate practice au impus-o, mai cu seamă în cazul anchetelor de opinie, procedură numită eșantionarea pe cote, care va face obiectul rîndurilor ce urmează.

Condiția de bază ce se cere îndeplinită pentru a putea realiza o eșantionare pe cote este cunoașterea prealabilă a cîtorva caracteristici de bază ale populației. In fond, eșantionarea pe cote poate fi considerată ca o eșantionare stratificată neprobabilistă. Cunoscînd distribuția populației în clasele caracteristicilor de stratificare se va calcula, din mărimea totală a eșantionului, cîți indivizi trebuie aleși din fiecare clasă a acestor caracteristici structurale. Dacă e vorba de o anchetă cu operatori de interviu, cum se întîmplă cel mai adesea, atunci fiecărui operator i se va indica, pe lîngă numărul total de persoane de anchetat, și însușirile pe care acestea trebuie să le posede, precum și modalitatea concretă de găsire și abordare a lor. Sarcinile ce revin unui singur operator se numesc cote, de unde și numele procedeului.

Să luăm și aici un exemplu, pentru a facilita înțelegerea tehnicii. Să zicem că ne interesează obținerea unui eșantion de 2.000 de studenți dintr-un centru universitar oarecare, spre a fi anchetați pe baza unui chestionar. Să mai presupunem că avem la dispoziție 100 operatori și deci că fiecare dintre aceștia este chemat să realizeze 20 de chestionare. Despre populația studențească a centrului universitar cunoaștem distribuția în funcție de:

- apartenanţa la cele patru institute de învăţămînt superior existente în oraş (notate cu A,B,C şi D):

A - 40% ; B - 30% ; C - 20% ; D - 10%

- sex:

băieţi - 55% ; fete 45%

- ani de studii:

I - 30% ; II - 25%; III - 20% ; IV - 15% ; V - 10%

Prin urmare, eşantionul nostru de 2000 de studenţi va trebui să conţină elemente astfel repartizate:

- după institute:

A - 800
B - 600
C - 400
D - 200

- după sex:

băieţi - 1.100
fete - 900

- după ani de studiu:

I - 600
II - 500
III - 400
IV - 300
V - 200

Presupunînd că sarcinile operatorilor sînt uniforme (adică identice pentru toţi), fiecare va trevui să chestioneze 20 de studenţi, repartizaţi după cum urmează:

- după institute:

A - 8; B - 6; C _ 4; D - 2

- după sex:

băieţi: - 11; fete - 9

- după ani de studiu:

I - 6; II - 5; III - 4; IV - 3; V - 2;

Acestea constituie deci cotele ce revin fiecăruia, urmînd ca în investigarea efectivă anchetatorul să-şi aleagă subiecţii pe care-i doreşte, cu condiţia ca în final să fie îndepli-

nite condiţiile de mai sus. Am ales aici, în scop didactic, un caz foarte simplu, dar cred că se înţelege că, în realitate, este ineficient să se repartizeze operatorilor sarcini identice, deoarece aceasta implică obligativitatea ca fiecare operator să meargă în toate cele patru institute. Sarcinile se pot concentra, în cazul de faţă, prin repartizarea a 40 de operatori exclusiv în institutul A, a 30 de operatori, în B, a 20 în C şi a 10 în D.

Criteriliile de sarcini de genul celor mai sus ilustrate se numesc independente, deoarece cei trei factori pe baza cărora se constituie cotele nu sînt legaţi unul de celălalt. Astfel, cei 4 studenţi din anul III pot fi aleşi tiţi dintr-un singur institut sau cîte unul din fiecare; ei pot fi toţi de acelaşi sex sau 2 băieţi şi 2 fete etc. In cazul cotelor independente nu se face deci nici o precizare prealabilă asupra raporturilor dintre cote. Desigur că de cele mai multe ori acest lucru poatea avea consecinţe neplăcute, în sensul că unele categorii vor apare suprareprezentate iar altele subreprezentate. De exemplu, în cazul nostru, dacă toţi operatorii vor raţiona astfel "am de ales 4 studenţi din anul III şi tot 4 din institutul C, atunci cel mai comod este ca mergînd în institutul C să caut direct anul III şi să-mi epuizez de aici sarcina", va rezulta în final că toate opiniile studenţilor de anul III vor fi caracteristice doar institutului C. Normal că am exagerat puţin aici şi că nu toţi vor proceda astfel, dar sînt şanse foarte mari ca unii să o facă şi atunci răspunsurile în eşantion vor fi distorsionate în aceste alegeri.

Pentru a evita asemenea situaţii, este preferabil ca să se lucreze cu cote legate, adică să se constituie, de fapt, un tabel de asociere simultană între toate caracteristicile avute în vedere. De pildă, un anchetator poate primi sarcina astfel:

	I N S T I T U T							
	A		B		C		D	
An / Sex	Băiat	Fată	Băiat	Fată	Băiat	Fată	Băiat	Fată
I	1	2	1	1	-	1	-	-
II	1	1	1	-	1	-	1	-
III	1	-	1	-	-	1	-	1
IV	1	1	-	-	1	-	-	-
V	-	-	1	1	-	-	-	-

iar ceilalţi în mod similar, în aşa fel încît pe ansamblu să iasă atît numărul dorit, dar şi anumite corelaţii între factorii de stratificare.

Este clar că nu putem utiliza criterii corelate decît dacă dispunem de o bază statistică după care să putem proiecta respectivele cote, adică atunci cînd ştim care sînt asocierile efective între caracteristici în populaţie.Chiar şi în cazul în care astfel de corelaţii sînt imprecise, se recomandă totuşi o formulare cît mai specifică a cotelor, pentru ca operatorul să aibă de a face cu cît mai multe elemente constrîngătoare în alegeri, să i se micşoreze pe cît posibil facilitatea de a înlocui un element cu altul. Desigur că nu trebuie nici exagerat pe această cale; dacă operatorul întîmpină prea multe obstacole în depistarea indivizilor cu caracterele indicate, el va avea tendinţa de a "fabrica" el însuşi asemenea personaje.

De asemenea, se mai recomandă, în anchetele de opinie bazate pe un eşantion pe cote, ca sarcinile operatorului să fie cît mai restrînse (numeric) cu putinţă , nedepăşind 15-2o de interviuri. Mai multe motive pledează pentru utilizarea unui număr mare de anchetatori cu sarcini reduse fiecare. Mai întîi, alegerile subiective, care nu pot fi totuşi înlăturate, se vor orienta spre un număr mai mare de direcţii, care, cu o probabilitate ridicată, vor duce la compensări reciproce. O sarcină mică reduce, de asemenea, ponderea pe întreg eşantionul a orientărilor subiective prea pronunţate ale unor operatori. Nu este tot una dacă un asemenea operator realizează 15 sau 15o de chestionare; efectul distorsionant este mult mai mare în cazul al doilea. Dacă au de făcut un număr prea mare de anchete, operatorii vor avea tendinţa de a le realiza cît mai operativ şi asta are efecte negative nu numai în modul de completare a chestionarului, ci şi în ceea ce priveşte grija de a selecta persoanele în eşantion. Cum spuneam, este important să se specifice anchetatorilor şi modul în care vor lua contact cu persoanele anchetate: acasă, la locul de muncă, pe stradă etc., în scopul de a evita contactarea prea numeroasă a unor tipuri specifice din populaţie. Or, este evident că un operator încărcat cu prea multe sarcini va alege, de preferinţă, cele mai accesibile mijloace de contactare a subiecţilor.

In literatura de specialitate, există multe discuţii referitoare la avantajele şi dezavantajele eşantionării pe cote (în general, a selecţiei nealeatoare), în raport cu cea aleatoare. La prima vedere, faptul că pentru aceasta din urmă este posibilă calcularea reprezentativităţii iar pentru prima nu, poate părea un avantaj decisiv, care să încline complet balanţa în favoarea studiilor pe eşantioane probabiliste. Iar dacă argumentul nu pare totuşi convingător se mai pot aduce şi altele în sprijin. E vorba, mai întîi de dificultăţile de a construi cote în raport cu cele mai relevante caracteristici pentru studiu, dat fiind că nu se cunoaşte, de cele mai multe ori, distribuţia lor în populaţie. In felul acesta este greu să se îngrădească subiectivitatea în alegere, subiectivitate care oricum rămîne în orice selecţie nealeatoare. In fine, un alt argument de marcă împotriva acestui gen de selecţie este dat de imposibilitatea realizării unui control eficient a muncii de teren a operatorilor. In eşantionul aleator, cel care conduce cercetarea cunoaşte identitatea fiecărei persoane ce trebuie anchetată şi deci poate merge pe urma operatorilor spre a verifica dacă şi cum s-au realizat anchetele. In elantionarea pe cote, mai cu seamă dacă chestionarul este anonim, pierdem orice posibilitate de control, nefiind siguri nici măcar dacă persoană chestionată există efectiv.

Cu toate acestea, studiile pe baza eşantioanelor nealeatoare nu numai că nu sînt rare, dar probabil, dacă s-ar face o statistică, s-ar constata că ele sînt întîlnite cu preponderenţă în sociologie şi psihologia socială. Lucrul se întîmplă, de regulă, prin alegerea deliberată a acestui gen de selecţie, dar mai există şi cazuri cînd cercetătorii au numai impresia sau lasă numai impresia că au ales un eşantion aleator, cînd de fapt, situaţia este cu totul alta.

Lăsînd la o parte acest din urmă caz, să vedem care sînt motivele principale ce duc totuşi la o largă utilizare conştientă a eşantioanelor nealeatoare, pe cote, în speţă. Trebuie să înţelegem, în prealabil, că factorul principal de care depinde modalitatea de efectuare a unei cercetări este eficienţa pe care dorim s-o aibă această investigaţie. Noţiunea de eficienţă înglobează, în primul rînd, rezultatele teoretice sau aplicative ale cercetării, rezultate ce depind în bună măsură de acu-

…ețea eșantionării în studiile selective. Dar alte laturi, …u pot fi neglijate, sînt costul cercetării și promptitudinea (rapiditatea) cu care se ajunge la rezultatele finale. Desigu că ponderea fiecărui factor depinde de genul de cercetare efe tuată, de scopul ei etc. In cazul unor cercetări cu caracter aplicativ, comandate de un beneficiar, ultimii doi factori ac ționează cu o forță deosebită.

Am făcut această paranteză pentru a înțelege mai bin de ce nu sînt rare situațiile cînd se preferă eșantioane pe c te. Intr-adevăr, costul realizării acestora este mult mai mic preparativele sînt relativ sumare și deci promptitudinea cerc tării este deosebită. Faptul e ușor de ilustrat prin investig țiile efectuate de institutele occidentale de sondare a opini publice asupra unor probleme precum: audiența emisiunilor rad și TV, preferințele în campaniile electorale etc. Este clar c asemenea informații trebuie culese și prelucrate rapid, altfel datele își pierd orice valoare. Or, numai un eșantion pe cot este capabil să furnizeze informația dorită în asemenea cazuri Cu siguranță că și eșantionarea pe cote presupune o fază de pr gătire atentă, bazată pe cunoașterea realității sociale inves tigate, fapt confirmat și de secretul în care este învăluită a ceastă etapă a investigației, în cazul institutelor de opinie publică occidentale. Intr-adevăr, procedura de alegere a unui eșantion puternic reprezentativ la nivel național pare să fie operația "tehnologică" din științele sociale al cărei secret este cel mai bine păzit. Experiența de pînă acum a dovedit că … că eșantionarea pe cote este minuțios preparată, datele obținu- te se pot extrapola la populația totală cu o marjă de eroare similară celei calculate pentru eșantioane probabiliste de acee talie. Astfel, ca să reiau exemplul institutelor menționate, în țări de mărimea Angliei, Franței sau R.F. a Germaniei, cu un e- șantion național mai mic de 2000 de persoane, studiat în preajm alegerilor, se pronosticheează rezultatele cu o precizie ce merge uneori pînă la mai puțin de 1%.

Asemenea performanțe sînt deseori subliniate de practi- cieni, în ciuda criticilor statisticienilor, pentru care singură selecția aleatoare rămîne capabilă să dea rezultate extrapola- bile. Tot practicienii mai remarcă, în mod judicios aș spune, că

de fapt eşantionul probabilistic ales nu este niciodată în intregime investigat (cînd e vorba de anchete, evident). Intr-adevăr există mereu un procent, ce deseori depăşeşte 30%, de indivizi care nu pot fi găsiţi acasă, sînt inapţi pentru răspunsuri, refuză să colaboreze etc. Chiar dacă numărul lor poate fi redus prin vizite repetate, este clar că fenomenul nu poate fi în întregime eliminat. E.Noelle (1966) aprecia că valoarea de 80% din eşantion poate fi greu atinsă, chiar dacă se fac pînă la 3 vizite. Apoi insistenţele pe lîngă persoanele care nu doresc să răspundă au mai degrabă un efect negativ asupra rezultatelor cercetării, iar cele pe lîngă persoanele inapte sînt cel puţin deplasate. Prin urmare, eşantionul, oricît de corect ar fi iniţial, şi mai ales dacă a fost corect ales, rămîne pînă la urmă trunchiat, fiind astfel purtătorul unor posibile erori sistematice.

Sînt, de asemenea, înclinat să dau dreptate celor care susţin că, de fapt, în anchetele pe bază de eşantion, erorile cele mai importante sînt departe de a fi cele datorate eşantionării, ele provenind mai degrabă din modul concret de administrare a chestionarului, din munca operatorilor, din felul în care se dau şi se înregistrează răspunsurile.

In fine, dar nu în ultimul rînd ca importanţă, ci poate chiar în primul, eşantionarea nonprobabilistică se impune din cauza inexistenţei unui cadru de eşantionare suficient de bun sau adus la zi şi a eforturilor enorme pe care un sociolog ar trebui să le facă pentru a-şi construi singur un astfel de cadru. Listele cu populaţia sînt de cele mai multe ori vechi şi incomplete, iar unele satisfăcătoare existente pot să nu-i fie accesibile sociologului.

Iată cîteva motive care explică şi pledează pentru folosirea pe scară largă a eşantioanelor nealeatoare. Expunerea lor şi sublinierea unor avantaje deloc neglijabile ale acestui gen de selecţie nu înseamnă o respingere de principiu a metodelor aleatoare. Dimpotrivă, gîndul oricărui sociolog trebuie îndreptat spre exploatarea tuturor posibilităţilor de a întocmi un eşantion probabilist. Numai dacă obstacolele ce stau în calea acestui plan sînt insurmontabile, în timpul şi cu mijloacele avute la dispoziţie, trebuie să apeleze, cu prudenţă şi vigilenţă, la folosirea unuia aleator.

2.3. Eşantioane fixe (panel)

Ieşind din firul logic ce a ghidat clasificarea folosi tă pînă aici (eşantionare aleatoare şi nealeatoare), voi pune, foarte pe scurt în evidenţă un tip deosebit de eşantioane, car nu se disting prin modul în care au fost construite, ci prin ce de utilizare a lor. Eşantioanele fixe (sau panel, după un terme englezesc) sînt acelea care, indiferent de modalitatea lor de alegere, servesc unor studii repetate la diferite intervale de timp.

In cazul în care ele sînt constituite din indivizi umani, ele ridică unele probleme specifice, ce se cer menţionate aici. Mai întîi, este limpede că repetarea unor cercetări, iden tice sau diferite ca problematică, asupra aceluiaşi eşantion aduce avantaje în ceea ce priveşte reducerea costului cercetări a timpului şi, dacă reprezentativitatea eşantionului s-a dovedit corespunzătoare, şi în privinţa acurateţei rezultatelor. Od tă ce indivizii au fost supuşi unei investigaţii, nu numai că e nu mai trebuie din nou selecţionaţi - cu toate riscurile şi gre tăţile caracteristice acestei operaţui - dar şi o serie de prob me legate de contactul cercetător-subiec angajarea subiectulu într-o discuţie deschisă etc. dispar.

Probabil însă că avantajul cel mai semnificativ al eşa tioanelor fixe se manifestă atunci cînd se repetă o anchetă, în scopul evaluării şi explicării schimbărilor sociale petrecute î intervalul de timp dintre cele două momente. Desigur, că şi stu diul pe două eşantioane diferite evidenţiază schimbările, dar atunci cînd eşantionul rămîne acelaşi se pot pune în evidenţă i divizii care şi-au modificat "comportamentul" în răstimpul dat, în raport cu o caracteristică studiată, şi din analiza situaţie lor se vor putea trage concluzii referitoare la factorii de schimbare.

Nu trebuie totuşi supraestimate virtuţile eşantioanelor panel. Multe din avantajele lor se transformă treptat în dezavantaje. Oamenii se pot plictisi de anchete repetate şi vor începe să refuze de a mai răspunde; unii mor, alţii se mută, alţi devin indisponibili pentru anchetă. Astfel, încet, eşantionul în

cepe treptat să se degradeze "fizic" şi trebuie mereu reîmprospătat, ceea ce evident îi anulează o parte din avantaje.

Pe lîngă degradarea "fizică", se poate vorbi şi de una "morală", în sensul că oamenii fie că vor deveni plictisiţi şi vor începe să răspundă tot mai superficial la întrebări, fie că vor suferi un fenomen de "condiţionare". Stiind că va fi interogat despre anume probleme, omul, de cele mai multe ori, îşi schimbă comportamentul faţă de factorii respectivi; şi nu numai comportamentul, dar şi atitudinile, opiniile, convingerile etc. De exemplu, ştiind că va fi intervievată periodic în legătură cu programele de televiziune, orice persoană va avea tendinţa să urmărească mai des decît în mod normal aceste emisiuni, să le judeţe cu ochiul celui care va trebui să le aprecieze calitatea în faţa cuiva etc. şi astfel este foarte probabil că organizatorul sondajului va constata rapid că în eşantionul său creşte timpul de vizionare a programelor şi se schimbă aprecierile acestora, cînd în realitate (adică în populaţie) lucrurile pot fi cu totul altele.

In ciuda celor de mai sus, eşantionarea panel rămîne o modalitate foarte eficientă de culegere rapidă a unor informaţii consistente, mai ales atunci cînd este vorba de a analiza factorii care influenţează schimbările în opiniile, atitudinile şi comportamentul indivizilor umani în raport cu anumite fenomene sociale care se derulează cu viteză ridicată şi nu este timp pentru punerea în aplicare a unor strategii de cercetare mai complexe, dar mai greoaie, mai puţin prompte în aplicare.

Cap. XII. ASPECTELE MATEMATICE ALE ESANTIONARII

1. Determinarea mărimii şi reprezentativităţii eşantionului. Estimarea mediilor şi proporţiilor.

In general, scopul studiului pe bază de eşantion îl constituie estimarea unor mărimi ale populaţiei, de regulă, medii sau proporţii. Spun estimare şi nu determinare, datorită faptului că prin intermediul unui eşantion nu putem obţine valoarea exactă urmărită - dacă în general are sens să vorbim de o valoare exactă - decît în cazuri cu totul excepţionale şi datorită unui hazard pe care nu ne putem prea mult baza. Dar pentru nevoile practicii este mai mult decît suficientă obţinerea unor valori estimative, apropiate de cea care admitem că există realmente în populaţie.

Să presupunem că am efectuat un studiu exhaustiv (asupra întregii populaţii) şi am obţinut următoarea caracteristică X:

Valori	x_1	x_2	x_3	...	x_s	Total
Frecvenţe	k_1	k_2	k_3	...	k_s	N

In acest caz se poate calcula media exactă a variabilei X:

$$\bar{x} = \frac{k_1x_1 + k_2x_2 + \ldots + k_sx_s}{N}$$

precum şi alte mărimi, precum abaterea standard σ_x. Efectuînd o cercetare selectivă, nu vom studiu pe toţi cei N indivizi din populaţie, ci numai o parte a lor, să zicem un număr n. Pentru aceştia obţinem o caracteristică X' de forma:

Valori	x_1	x_2	x_3	...	x_s	Total
Frecvenţe	k'_1	k'_2	k'_3	...	k'_s	n

[illegible]a medie va fi:

$$\bar{x}' = \frac{k_1'x_1 + k_2'x_2 + \ldots + k_s'x_s}{n}$$

aproape sigur diferită de $\bar{x}$.

Diferenţa $\bar{x} - \bar{x}'$ se numeşte eroarea introdusă de selecţia care a condus la mărimea $\bar{x}'$. Este evident că pentru orice cercetare este preferabil să se obţină o eroare cît mai mică, adică cele două medii să fie cît mai apropiate. Dar tot atît de evident este că dacă eroarea nu depăşeşte o anumită limită, estimarea lui $\bar{x}$ prin $\bar{x}'$ este absolut satisfăcătoare pentru scopurile cercetării şi depunerea unor eforturi suplimentare pentru a reduce diferenţa $\bar{x} - \bar{x}'$ nu este rentabilă, deoarece se măresc cheltuielile şi timpul reclamate de investigaţie.

De pildă, dacă urmărim evaluarea cîtşigului mediu lunar realizat de o colectivitate de muncitori, în unele cazuri este suficient să ştim că eroarea comisă nu depăşeşte 5o lei. Aceasta înseamnă că dacă obţinem în eşantion un cîştig mediu de 22oo lei, vom şti că cel care caracterizează populaţia va fi de această mărime $\pm$ 5o lei. In mod identic se pune problema evaluării unei proporţii. Dacă ne satisface o eroare de 1%, atunci prin cercetarea selectivă se va urmări respectarea condiţiilor ca proporţia obţinută să nu difere de cea "adevărată" cu mai mult de un procent: $|p - p'| < 0{,}01$. De exemplu, dacă în eşantion găsim o proporţie de activi de o,45(45%), atunci în populaţie ea va fi de o,45 $\pm$ o,ol, deci între 44% şi 46%.

De fapt, afirmaţiile de mai sus trebuie - în aparenţă puţin dar în realitare fundamental - corijate. Dat fiind că valoarea adevărată din populaţie (media $\bar{x}$ sau proporţia p) este necunoscută (altfel n-ar avea rost studiul), nu se va putea afirma niciodată <u>cu certitudine</u> că diferenţa $|\bar{x} - \bar{x}'|$ sau $|p - p'|$ este mai mică decît o anumită valoare. Teoria matematică a probabilităţilor ne ajută la stabilirea unui alt gen de afirmaţie, şi anume la a putea spune că respectiva diferenţă este mai mică decît o mărime Δ aleasă după exigenţele de precizie ale cercetării, numită <u>eroare maximă admisă</u>, <u>cu o anumită probabilitate P</u>, suficient de ridicată pentru a o putea considera ca o cvasicertitudine. Mai scriem această probabilitate, numită şi <u>nivel de încredere</u>, sub forma $P(|\bar{x} - \bar{x}'| < \Delta)$ şi o citim: "probabilitatea ca me-

dia din populație și cea din eșantion să fie mai mică decît Δ ". Nivelul de încredere cel mai jos utilizat este de 95% (sau o,95) iar în pracrică, pentru cazuri de exigență sporită, se mai folosesc cele de 99% (o,99) și 99,9% (o,999).

Se vede deci că <u>precizia</u> oricărei eșantionări privește două elemente: eroarea Δ și probabilitatea P. Ambele elemente de incertitudine , Δ și P, se datorează faptului că eșantionul este doar o parte a populației, al doilea element fiind, în același timp, determinat și de necunoașterea mărimii pe care vrem s-o estimăm. Cuplul (Δ , P) constituie ceea ce, de obicei, se numește <u>reprezentativitatea</u> eșantionului. Deci reprezentativitatea privește în același timp intervalul în jurul mediei din eșantion $(\bar{x}'-\Delta, \bar{x}'+\Delta)$ - numit și <u>interval de confidență</u> - cît și probabilitatea cu care putem afirma că media din populație, $\bar{x}$, cade în acest interval.

Vom spune că un eșantion A este mai reprezentativ decît unul B dacă, la același interval de confidență, nivelul de probabilitate al lui A este mai ridicat decît cel al lui B sau dacă, la nivele egale de încredere, intervalul de confidență al lui A este mai mic sau dacă, simultan, nivelul de încredere e mai mare și intervalul mai mic. Un eșantion poate fi numit pur și simplu reprezentativ (deci fără comparație cu altul) dacă cele două mărimi se încadrează în limitele cerute de necesitățile cercetării, adică Δ este mai mic decît eroarea maximă admisă iar P mai mare decît gradul minim de probabilitate admis: o,95.

Este clar, din afirmația făcută puțin mai sus asupra surselor incertitudinii, că mărimea gradului de reprezentativitate al eșantionului se poate realiza, în condiții egale, prin sporirea volumului eșantionului. Atunci cînd <u>n</u> se apropie de N, Δ va tinde spre 0 (exactitudine) iar P spre 1 (certitudine). Deci este o funcție descrescătoare de <u>n</u> iar P una crescătoare. Dar la un <u>n</u> fixat mărimile Δ și P nu sînt independente una de alta. Reducînd eroarea Δ (deci sporind exactitatea estimării) se micșorează probabilitatea P(adică șansele ca această exactitate să se realizeze) și invers, mărind probabilitatea, mărim și eroarea. Este, dacă vreți, o analogie, evident formală, cu celebrul principiu al incertitudinii al lui Heisenberg.

In ceea ce priveşte reprezentivitatea eşantionului, deci mărimile Δ şi P, se pot pune două probleme:

1) Să se determine mărimea n a eşantionului astfel încît să se realizeze reprezentativitatea aleasă (un Δ şi un P). De exemplu, dacă reluăm cazul ilustrat mai înainte, să se determine numărul muncitorilor ce trebuie anchetaţi, pentru a putea spune cu o probabilitate de, să zicem, 95% (P = o,95) că media din populaţie nu diferă cu mai mult de 5o lei de cea din eşantion (Δ = 5o lei).

2) Avînd dat un eşantion de mărime n, să se aprecieze reprezentativitatea acestuia, deci să se calculeze eroarea maximă la un anumit nivel de probabilitate. Adică, dacă am investigat 1ooo de muncitori şi am obţinut o medie de venituri de 22oo lei, să se vadă la nivelul de probabilitate de o,95 (sau altul mai mare) în ce interval de confidenţă cade media din populaţie.

Prima problemă se ridică în faza de proiectare a cercetării iar a doua în cea de prelucrare a datelor. Aşa cu, se va vedea mai jos, este greşit să se creadă că rezolvarea primeia implică şi rezolvarea secundeia, calculele trebuind făcute şi înainte şi după efectuarea investigaţiei concrete. De asemenea, trebuie subliniat că anumite formulări, din păcate destul de des întîlnite în studiile sociologice de teren, de genul "s-a studiat problema cutare pe un elantion reprezentativ" sînt absolut fără nici o semnificaţie dacă nu se precizează cei doi indicatori ai reprezentativităţii: Δ şi P.

Problemele propuse se rezolvă în felul următor. Să ne imaginăm că am studiat un eşantion A de volum n şi am obţinut o medie $\bar{x}'_a$. Extragem apoi un alt eşantion B, de acelaşi volum, şi obţinem o medie $\bar{x}'_b$ - Repetînd indefinit operaţia, se va obţine un şir de valori: $\bar{x}'_a$, $\bar{x}'_b$, $\bar{x}'_c$ ...., care, fiecare, constituie o estimaţie a valorii din populaţie $\bar{x}$. Se ajunge astfel la o variabilă statistică cu valorile de mai sus, care, la rîndul său, va avea o medie (media tuturor mediilor eşantioanelor), notată cu $\bar{\bar{x}}'$ şi o abatere standard $\sigma_{\bar{x}'}$. Primul indicator arată în jurul cărei valori se grupează mediile diferitelor eşantioane iar ultimul, gradul de împrăştiere faţă de poziţia centrală.

Teoria matematică a probabilităţilor demonstrează,

pentru variabila construită mai sus, următoarele trei propoziţii fundamentele, necesare şi suficiente pentru rezolvarea problemele enunţate:

a) Valoarea medie a distribuţiei mediilor eşantioanelor este egală cu media populaţiei:

$$\bar{\bar{x}}' = \bar{x} \qquad (1)$$

Adică luînd valorile medii obţinute în fiecare eşantion în parte şi făcînd media acestora, se ajunge exact la valoarea căutată di populaţie.

b) Varianţa (pătratul abaterii standard) distribuţiei mediilor eşantioanelor este de n ori mai mică decît varianţa distribuţiei indivizilor în populaţie:

$$\sigma^2_{\bar{x}'} = \frac{\sigma^2_x}{n} \qquad (2)$$

Se observă deci că împrăştierea mediilor eşantioanelor în jurul lui $\bar{x}$ este relativ mică şi depinde de dispersia variabilei studiate (directe proporţional) şi de volumul eşantionului (invers proporţional). Este prin urmare evident că mărirea numărului de indivizi în eşantion conduce la reducerea împrăştierii, deci la o "strîngere" a mediilor eşantioanelor în jurul mediei populaţie

Formula de mai sus este, strict vorbind, valabilă numai în cazul selecţiei simple aleatoare repetate. Pentru selecţi nerepetată, intervine un factor de corecţie ce modifică formula astfel:

$$\sigma^2_{\bar{x}'} = \frac{\sigma^2_x}{n} \cdot \frac{N - n}{N - 1}$$

unde N este volumul populaţiei. Pentru a nu complica prea mult expunerea , în cele ce urmează se va folosi doar formula mai simplă (2).

c) Variabila statistică $\bar{x}'$, adică mediile eşantioanelor, se distribuie după o curbă normală.

Cîteva cuvinte explicative se impun în legătură cu curba normală. Este vorba de o funcţie matematică dată de formula:

$$f(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u-\bar{u}}{\sigma}\right)^2}$$

unde $\underline{u}$ este variabila independentă, $\bar{u}$ - valoarea medie, iar σ abaterea standard; $\underline{e}$ fiind cunoscuta constantă matematică, baza logaritmilor naturali. (Nu am notat cu $\underline{x}$, cum este obiceiul, variabila independentă pentru a nu o confunda cu variabila urmărită în cercetarea noastră, despre care nu putem presupune că ar fi normală). Graficul funcţiei f(u) este cunoscuta curbă a lui Gauss, sub formă de clopot, simetrică în raport cu valoarea medie şi tinzînd asimptotic la zero, spre $-\infty$ şi $+\infty$. Forma mai plată sau mai ascuţită a curbei depinde de celălalt parametru care apare în formulă: abaterea standard. Cu cît ea este mai mare, cu atît curba este mai plată şi se apropie mai greu de axa absciselor (fig.1).

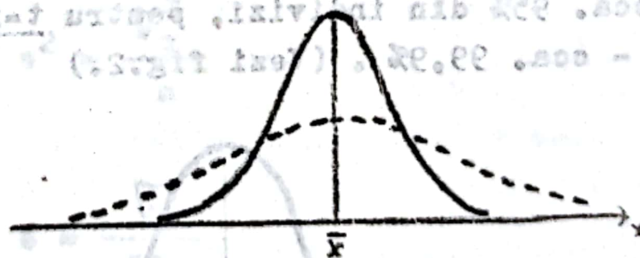


Fig.1. Două curbe normale, cu aceeaşi medie şi abateri standard diferite.

Ca orice funcţie de distribuţie, şi aceasta are proprietatea că aria determinată de ea şi axa absciselor, între două puncte oarecare u_1 şi u_2 adică integrala funcţiei de la u_1 la u_2 dă probabilitatea ca un individ statistic oarecare să ia o valoare cuprinsă între u_1 şi u_2. Deci:

$$P(u_1 \leq u \leq u_2) = \int_{u_1}^{u_2} f(u)du$$

In cazul funcţiei normale, dacă valorile u_1 şi u_2 sînt simetrice în raport cu media, adică formează un interval al cărui punct central este chiar media, se pot găsi probabilităţile respective cu ajutorul unor tabele, în funcţie de lungime intervalului respectiv. Aceste probabilităţi depind de forma curbei, adică de abaterea standard, dar valorile vor deveni independente de respectivul parametru, dacă lungimea intervalului este măsurată tocmai în abateri standard. Valorile u_1 şi u_2 sînt de forma:

$$u_1 = \bar{u} - t\sigma$$

$$u_2 = \bar{u} + t\sigma$$

unde t este un număr care arată cu cîte abateri standard capetele intervalului se află la stînga şi la dreapta mediei. Deci curbele normale au remarcabila proprietate că toate conţin, într-un interval de acest gen, aceeaşi proporţie de indivizi. Astfel se constată că pentru $t=2$, deci pentru un interval de două abateri standard în stînga şi dreapta mediei, corespunde o putere de cuprindere de cca. 95% din indivizi, pentru $t=2,6$ - cca. 99% şi pentru $t = 3,3$ - cca. 99,9% . (Vezi fig.2.)

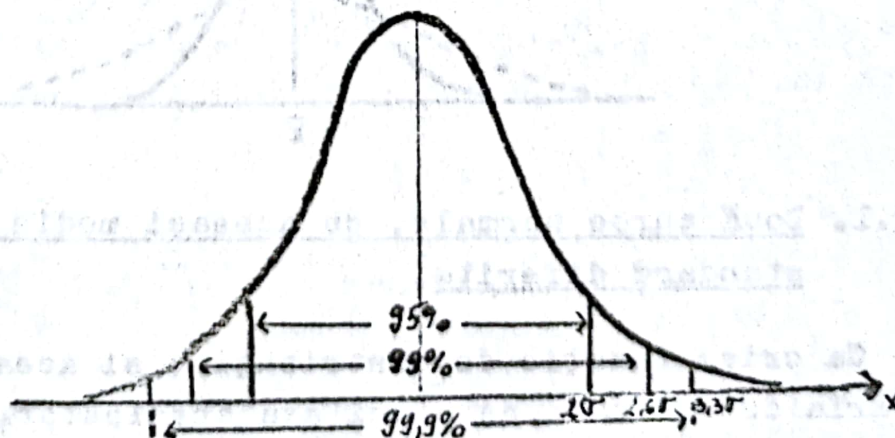


Fig.2. Procentul din aria unei curbe normale

Altfel spus, pentru orice distribuţie normală, există 95% şanse ca un individ să se abată de la medie cu mai puţin de două abateri standard, 99% să se abată cu mai puţin de 2,6σ şi 99,9% cu mai puţin de 3,3σ.

Să revenim acum la eşantioanele noastre imaginare. Mediile lor urmează deci o curbă normală, în jurul mediei populaţiei, pe care fiecare medie de eşantion o estimează. Noi nu vom şti niciodată exact cît de bună este estimaţia realizată cu un eşantion anume, cel pe care-l folosim de fapt în cercetare. Ştim însă acum cîteva lucruri esenţiale. Mai întîi, formula (2) ne dă varianţa mediilor de eşantioane. Abaterea standard respectivă arată gradul de precizie al eşantionării folosite, reflectînd eroarea cea mai probabilă, motiv pentru care poartă denumirea de <u>eroare standard</u>. O vom nota, din motive de comoditate a scrierii cu <u>e</u>. In al doilea rînd, din consideraţiile anterioare rezultă că există 95% şanse ca un eşantion oarecare să producă o eroare mai mică decît două erori standard (2e), 99% şanse ca eroarea să fie mai mică decît 2,6e şi 99,9% şanse ca ea să nu depăşească 3,3e. (Se observă că valorile alese pentru <u>t</u>: 2; 2,6 şi 3,3 s-au luat pentru a cădea peste nivelele de probabilitate utilizate mai des).

Cele două probleme enunţate se rezolvă acum foarte simplu. Observăm că formula (2) devine, după schimbarea notaţiei:

$$e^2 = \frac{\sigma_x^2}{n} \qquad (3)$$

sau

$$e = \frac{\sigma_x}{\sqrt{n}} \qquad (4)$$

A. <u>Determinarea mărimii eşantionului</u>

Din formula (3) se scoate <u>n</u> şi se obţine:

$$n = \frac{\sigma_x^2}{e^2} \qquad (5)$$

Mărimea <u>e</u> se determină în funcţie de exigenţele cercetării. Se stabileşte mai întîi eroarea maximă admisă, Δ, care co-

roborată cu un nivel de probabilitate conduce la o valoare pentru <u>e</u>. In adevăr, dacă alegem $P = 0,95$, atunci $\Delta = 2e$ şi deci $e = \Delta/2$.

Ceea ce nu cunoaşte, însă din formula (5) este valoarea lui σ_x, adică abaterea standard în populaţie. In etapa iniţială a cercetării, această mărime se evaluează, mai exact se înlocuieşte cu o valoare despre care să fim siguri că nu va fi depăşită în realitate. Aşa după cum se vede din formula (5), cu cît vom da lui σ_x o valoare mai mare cu atît volumul eşantionului se va mări şi precizia va spori şi ea. Deci singura grijă ce trebuie avută e să nu subestimăm pe σ_x.

<u>Exemplu.</u> Urmărim să estimăm cu un eşantion vîrsta medie a unei colectivităţi de studenţi, a căror vîrstă ştim că variază aproximativ între 19 şi 25 ani. Imprăştierea cea mai mare s-ar obţine atunci cînd jumătate dintre indivizi ar fi de 19 ani iar ceilalţi de 25 de ani, fiind atunci de 22 ani. S-ar ajunge la o abatere standard de 3 ani (22 - 19 sau 25 - 22), mărime pe care o vom folosi în determinarea lui <u>n</u>. Drept eroare maximă admisă vom alege 3 luni, adică un sfert de an:

$$\Delta = 0,25$$

La nivelul de probabilitate de 0,95 această mărime conţine două erori standard. Deci:

$$e = 0,25 : 2 = 0,125$$

Atunci:

$$n = \frac{3^2}{0,125^2} = \frac{9}{0,0156} = 576$$

cifră care se poate rotunji la 6oo pentru mai multă siguranţă.

B. <u>Determinarea reprezentativităţii eşantionului</u>

Prin determinarea reprezentativităţii eşantionului înţelegem, aşa cum am spus, calculul erorii maxime admise, Δ, şi alegem nivelului de probabilitate P. Acest lucru se face plecînd de la valoarea erorii standard, <u>e</u>, dată de formula (4). Dar nici acum la sfîrşitul cercetării nu sîntem în posesia valorii exacte σ_x. In schimb, cunoaştem abaterea standard a variabilei X' în

cadrul eşantionului studiat: $\sigma_{x'}$. Teoria statisticii matematice arată că aceasta reprezintă o estimare suficient de bună a lui $\sigma_{x'}$, putînd fi folosită în formulă. Este clar că se ajunge doar la o estimare a erorii standard, însă, pentru simplificarea limbajului, vom considera, în cele ce urmează, că am obţinut chiar eroarea standard.

In rezumat: se determină eroarea standard e, se alege un nivel de probabilitate P, căruia îi corespunde un număr t, ce înmulţit cu e, va conduce la Δ, deci la un interval de confidenţă.

Să reluăm exemplul anterior. Să presupunem că am înregistrat vîrsta celor 600 de studenţi anchetaţi, obţinînd o medie:

$$\bar{x}' = 22,5 \text{ ani}$$

şi o abatere standard:

$$\sigma_{x'} = 1,5 \text{ ani.}$$

Eroarea standard va fi:

$$e = \frac{1,5}{\sqrt{600}} = 0,06$$

Pentru diferite nivele de probabilitate, vom avea:

$P = 0,95 \quad \Delta = 2e = 2.0,06 = 0,1$ ani
$P = 0,99 \quad \Delta = 2,6e = 2,6.0,06 \approx 0,16$ ani
$P = 0,999 \quad \Delta = 3,3e = 3,3.0,06 \approx 0,2$ ani

Prin urmare, sînt 95% şanse ca vîrsta medie a studenţilor să fie cuprinsă între 22,5 - 0,12 ani şi 22,5 + 0,12 ani, adică în intervalul (22,38, 22, 62), 99% în intervalul mai larg (22, 34, 22, 66) şi 99,9% în intervalul şi mai mare (22,3, 22,7).

Un alt exemplu. Să se determine volumul eşantionului necesar pentru estimarea mediei notelor obţinute de studenţi la o sesiune de examene, cu o precizie de 1/10 de notă.

Avem deci $\Delta = 0,1$. Dacă lucrăm cu nivelul de probabilitate de 0,95, atunci $\Delta = 2e$ şi, în consecinţă:

$$e = 0,05$$

Notele studenţilor sînt cuprinse practic între 4 şi 1o. Abaterea standard maximă poate fi 3, corespunzînd unei împărţiri a studenţilor în două clase egale: unii de nota 4, alţii de 1o. Stim însă că o asemenea situaţie este absurdă, putînd reduce, de la bun început şi fără nici o teamă, cifra la 2. Atunci:

$$n = \frac{\sigma_x^2}{e^2} = \frac{2^2}{0{,}05^2} = \frac{4}{0{,}0025} = \frac{40000}{25} = 1600$$

Studiind un eşantion de 16oo studenţi, să presupunem că am obţinut o medie a notelor:

$$\bar{x}' = 7{,}80$$

şi o abatere standard în eşantion:

$$\sigma_{x'} = 1{,}6 \text{ note}$$

Cu această din urmă valoare estimăm eroarea standard:

$$e = \frac{1{,}6}{\sqrt{1600}} = 0{,}04$$

In situaţia de faţă, putem afirma că sînt 95% şanse ca media reală a notelor să fie cuprinsă între 7,8o - 2.o,o4 şi 7,8o + 2.o,o4, adică în intervalul (7,72, 7,88).

Estimarea proporţiilor

Din punct de vedere formal, problemele de estimare a proporţiilor prin eşantionare se pot reduce la cele similare pentru medii. Intr-adevăr, avînd un atribut A, vom acorda valoarea 1 tututuror indivizilor care posedă atributul respectiv şi 0 celor care nu-l posedă. Se formează atunci în populaţie următoarea caracteristică:

Valori	1	0	Total
Precvențe	k_1	k_2	N

Să notăm:

$$p = \frac{k_1}{N}, \quad q = \frac{k_2}{N}$$

unde p reprezintă proporția celor care posedă atributul A și q a celor care nu-l posedă, cu proprietatea evidentă că $p + q = 1$.

Media caracteristicii va fi:

$$\bar{x} = \frac{k_1 \cdot 1 + k_2 \cdot 0}{N} = \frac{k_1}{N} = p$$

Deci media caracteristicii este tocmai p sau, altfel spus, caracteristica a fost construită de așa manieră încît proporția p să devină o medie. Se poate calcula și abaterea standard:

$$\sigma_x = \sqrt{\frac{(1-p)^2 k_1 + (0-p)^2 k_2}{N}} = \sqrt{(1-p)^2 \frac{k_1}{N} + p^2 \frac{k_2}{N}} =$$

$$= \sqrt{(1-p)^2 p + p^2 q}$$

Dar fiindcă

$$1 - p = q$$

$$\sigma_x = \sqrt{q^2 p + p^2 q} = \sqrt{pq(q+p)} = \sqrt{pq}$$

Deci

$$\sigma_x = \sqrt{pq} = \sqrt{p(1-p)}$$

Așadar, orice proporție p poate fi interpretată ca valoare medie a unei variabile cu valorile 1 și 0 și abatere standard dată de formula de mai sus. Fiindcă în problemele de estimare se pune, cum s-a văzut, problema găsirii unei valori maximale a abaterii standard, vom menționa că, pentru cazul de față, se poate demonstra simplu că σ_x atinge valoarea maximă atunci cînd $p=0,5$,

deci cînd jumătate din indivizi posedă atributul respectiv şi jumătate nu-l posedă. Această valoare maximă va fi:

$$\sigma_x = \sqrt{o,5(1-o,5)} = \sqrt{o,5.o,5} = o,5$$

Cu aceste precizări putem trece la prezentarea unui exemplu de estimare a proporţiilor. Să zice, că ne interesează aflarea procentului de indivizi dintr-o populaţie, care împărtăşesc o anumită opinie O. Fie p proporţia în populaţie şi p' cea obţinută în eşantion. Dorim ca diferenţa dintre p şi p' să nu depăşească 2%, adică, în limbajul proporţiilor, o,o2. Deci:

$$\Delta = o,o2 \qquad e = o,ol \qquad \text{pentru } P = o,95$$

Atunci volumul eşantionului va fi:

$$n = \frac{o,5^2}{o,ol^2} = \frac{o,25}{o,oool} = 25oo$$

Să presupunem în continuare că ancheta desfăşurată pe eşantionul de 25oo indivizi conduce la rezultatul că 2o% dintre aceştia împărtăşesc opinia respectivă. Deci:

$$p' = o,2 \qquad \text{şi} \qquad q' = o,8$$

Abaterea standard în eşantion va fi:

$$\sigma_{x'} = \sqrt{p'q'} = \sqrt{o,2.o,8} = \sqrt{o,16} = o,4$$

Cu ajutorul ei estimăm eroarea standard:

$$e = \frac{o,4}{\sqrt{25oo}} = \frac{o,4}{5o} = o,oo8$$

Pentru $P = o,95$, $\Delta = 2e = 2.\ o,oo8 = o,o16$ sau 1,6%.

In consecinţă, sînt 95% şanse ca procentul indivizilor din populaţie care împărtăşesc opinia O să fie cuprins între 2o-1,6 = 18,4% şi 2o +1,6 = 21,6%. Dacă dorim să ridicăm nivelul de încredere la 99% , atunci eroarea se va ridica şi ea la aproximativ 2% (2,6x o,oo8) şi deci intervalul de confidenţă va fi (18%, 22%).

2. Teste de semnificaţie

2.1. Semnificaţia diferenţei mediilor sau proporţiilor

De multe ori în lucrările sociologice întîlnim expresii de genul: "diferenţa dintre valorile a şi b este semnificativă" sau "colectivitatea A diferă semnificativ de B" etc- Trebuie făcută următoarea precizare esenţială: Se poate vorbi de semnificaţia diferenţei dintre două valori numai în cazul în care cel puţin una este rezultatul unui studiu pe bază de eşantion. Dacă mărimile respective sînt calculate pe populaţii întregi, pe care le reprezintă, nu are nici un sens de a vorbi despre semnificaţia lor (în sensul statistic al acestui termen, evident). Aşa de pildă, dacă vîrsta medie a unei colectivităţi A este de 45 ani, iar a uneia B de 46 ani, diferenţa de un an este diferenţa reală şi nu are nici un rost să i se aplice calificativul de semnificativă sau nesemnificativă. Dimpotrivă, dacă cele două populaţii sînt investigate numai selectiv, problema devine reală datorită faptului, cunoscut de noi de pe acum, că prin atari studii se obţin numai evaluări ale mediilor reale, evaluări incluse în intervale de posibilă "fluctuaţie". Or, este clar că anumite diferenţe pot apare numai datorită acestor fluctuaţii normale oricărei eşantionări. Dimpotrivă, dacă diferenţele vor depăşi anumite limite, ele nu mai pot fi puse pe seama impreciziei eşantionului, ci trebuie să reflecte diferenţe reale în populaţiile mamă din care eşantioanele provin. Aceste scurte explicaţii se vor clarifica şi preciza în analiza concretă a unor cazuri simple .

A. Primul tip de analiză priveşte situaţia cînd una dintre cele două valori care se compară provine dintr-o cercetare selectivă iar cealaltă nu. Să presupunem că este vorba de o medie $\bar{x}'$, ce trebuie comparată cu una din populaţie $\bar{x}$. Cum se ştie, dacă diferenţa dintre valori este mai mică decît două erori standard, ea poate fi pusă pe seama flucturaţiilor de eşantionare. Algebric se scrie:

$$|\bar{x} - \bar{x}'| < 2e$$

sau

$$\frac{|\bar{x} - \bar{x}'|}{e} < 2$$

In general, se calculează mărimea din stînga semnului de inegalitate, notată cu z

$$z = \frac{|\bar{x} - \bar{x}'|}{e}$$

şi se testează aşa-numita ipoteză nulă: "nu există diferenţă semnificativă între $\bar{x}$ şi $\bar{x}'$". Vom avea mai multe situaţii:

- Dacă $z < 2$, adică ne aflăm sub pragul de probabilitate de 0,95, ipoteza nulă nu poate fi respinsă, deci nu se poate afirma că există o diferenţă semnificativă între $\bar{x}$ şi $\bar{x}'$ (Atenţie! Nu faptul că diferenţa este nesemnificativă, ci că ea nu este semnificativă)

- Dacă $z > 2$, înseamnă că ipoteza nulă poate fi respinsă - diferenţa este semnificativă - cu 95% şanse.

- Dacă z este mai mare decît valori superioare lui 2, atunci gradul de întemeiere a respingerii ipotezei nule este mai ridicat. De pildă, dacă z depăşeşte valoare 3.3, respingerea ipotezei nule este aproape o certitudine şi despre diferenţă se spune uneori că este foarte semnificativă, deşi calificativul de "foarte" are aici doar rolul de a sugera depăşirea unui înalt grad de probabilitate.

Observaţie importantă. In limbajul statisticii, mai ales cînd e vorba de testele de semnificaţie, s-a introdus obişnuinţa ca nivelele de probabilitate utilizate, de exemplu 0,95; 0,99; 0,999 etc. să fie notate prin diferenţele acestor probabilităţi la unitate, adică: 0,05; 0,01; 0,001; etc. Astfel se spune că diferenţa $\bar{x} - \bar{x}'$ este semnificativă la nivelul 0,05. Asta înseamnă pur şi simplu că sînt 95% şanse ca ea să fie semnificativă.

Exemple.

a) Cu ajutorul acestui tip de analiză putem verifica reprezentativitatea eşantionului ales, dacă cunoaştem un parametru din populaţie. Voi prezenta un caz de acest fel furnizat de lucrarea "Studentul şi societatea" (1973) scrisă pe baza rezultatelor unei anchete pe un eşantion de 2172 studenţi. In cadrul eşantionului a fost găsită o proporţie de 54,9% băieţi, ştiindu-se în acelaşi timp, că procentul real al studenţilor de sex masculin pe ţară era, la data anchetei, de 55,6%. Intrebarea care se ridică este dacă diferenţa de 0,7% se poate datora unor fluctuaţii normale

de eşantionare sau există un factor sistematic care a distorsionat eşantionul. In cazul în care se constată - şi se va vedea imediat că într-adevăr se constată - că diferenţa nu este semnificativă, putem transfera cu destulă siguranţă acest calificativ de corectitudine a eşantionării şi asupra altor caracteristici investigate.

Cu notaţiile noastre vom avea:

$p = 0,556$ (proporţia în populaţie)

$p = 0,549$ (proporţia în eşantion)

$$\sigma_{x'} = \sqrt{0,556(1-0,556)} = 0,497$$

$$e = \frac{0,497}{\sqrt{2172}} = 0,0107$$

$$|p - p'| = 0,556 - 0,549 = 0,007$$

$$z = \frac{|p-p'|}{e} = \frac{0,0070}{0,0107} = 0,65$$

Prin urmare, z este mai mic decît 2 şi diferenţa nu este semnificativă.

b) Să presupunem că media de vîrstă pentru muncitorii dintr-o ramură industrială este de 31 ani. Pe un eşantion de 625 de muncitori din ramura respectivă, <u>în cadrul unui judeţ A,</u> s-a obţinut o medie de 32 de ani şi o abatere standard de 5 ani. Intrebarea este dacă putem afirma că în judeţul A media de vîrstă este mai ridicată decît pe ansamblul ţării sau diferenţa de 1 an poate fi pusă pe seama unor fluctuaţii normale de eşantionare.

Avem deci:

$$\bar{x} = 31; \quad \bar{x}' = 32; \quad \sigma_{x'} = 5$$

Va rezulta:

$$e = \frac{5}{\sqrt{625}} = \frac{5}{25} = 0,2$$

$$z = \frac{|\bar{x} - \bar{x}'|}{e} = \frac{1}{0,2} = \frac{10}{2} = 5$$

Deci z este mai mare chiar decît 3,3 şi putem spune că diferenţa este semnificativă la nivelul de probabilitate o,ool (adică sînt 99,9% şanse ca vîrsta medie în judeţul A să fie mai mare decît pe ţară.

B. Al doilea tip de analiză are în vedere compararea valorilor medii (proporţiilor) din două eşantioane diferite(chiar şi ca mărime). Avînd două populaţii A_1 şi A_2, presupunem că am efectuat o cercetare pe cîte un eşantion n_1 (din A_1) şi n_2 (din A_2), obţinînd pentru o aceeaşi variabilă mediile $\bar{x}'_1$ şi $\bar{x}'_2$, cu abaterile standard $\sigma_{x'_1}$ şi $\sigma_{x'_2}$.

Raţionînd ca în cazul unei singure populaţii, putem şi aici imagina o infinitate de cupluri de eşantioane, obţinînd pentru fiecare cuplu o variabilă dată de diferenţa mediilor eşantioanelor.

Teoria statisticii demonstrează că pentru această variabilă eroarea standard este dată de formula:

$$e = \sqrt{\frac{\sigma^2_{x'_1}}{n_1} + \frac{\sigma^2_{x'_2}}{n_2}}$$

Exemplu. Fie A_1 poppulaţia studenţilor de la Universitate şi A_2 de la Institutul Agronomic. Să admitem următoarele date obţinute prin sondaj:

$$n_1 = 9oo \qquad n_2 = 7oo$$

$$\bar{x}' = 22 \text{ ani} \qquad \bar{x}' = 22,5 \text{ ani}$$

$$\sigma_{x'_1} = 1,5 \text{ ani} \qquad \sigma_{x'_2} = 1,6 \text{ ani}$$

Atunci

$$e = \sqrt{\frac{1,5^2}{9oo} + \frac{1,6^2}{7oo}} = o,o8$$

$$z = \frac{|22,5 - 22|}{o,o8} = \frac{o,5}{o,o8} = \frac{5o}{8} = 6,3$$

Valoarea z depăşeşte şi 3,3 deci diferenţa este semnificativă.

Cu alte cuvinte se poate afirma că la Agronomie media de vîrstă este mai ridicată decît la Universitate.

2.2. Testul χ^2

Se folosește mai ales pentru caracteristicile calitative, spre a se stabili semnificația diferențelor între frecvențe. Se pot imagina multe situații de aplicare a testului, dintre care mai frecvente sînt următoarele: a) compararea frecvențelor diferitelor clase ale unei aceleași caracteristici; b) compararea distribuțiilor a două caracteristici cu același număr de clase; c) stabilirea existenței asocierii într-un tabel de asociere.

In esență, se pornește de la anumite tabele ce cuprind un număr de căsuțe cu frecvențele respective. Să presupunem că avem s căsuțe cu frecvențele empirice $n_1, n_2, \ldots, n_s$. Aceste frecvențe se vom compara cu un șir corespunzător de frecvențe numite teoretice (numite așa pentru motivul că adesea ele se deduc dintr-o ipoteză oarecare și nu sînt culese din realitatea empirică): $m_1, m_2, \ldots, m_s$.

Mărimea χ^2 e dată de formula:

$$\chi^2 = \frac{(n_1 - m_1)^2}{m_1} + \frac{(n_2 - m_2)^2}{m_2} + \ldots + \frac{(n_s - m_s)^2}{m_s}$$

χ^2 este o variabilă statistică pentru care se pot calcula probabilitățile de a obține valori mai mari decît anumite valori date. Mai exact, pentru anumite praguri de probabilitate: o,o5, o,ol etc. se pot determina valorile χ_o^2, valori care nu sînt depășite decît cu probabilitățile respective (ce pot fi oricît de mici). Aceste valori critice, care se găsesc în tabele special întocmite, depind nu numai de probabilități ci și de un alt parametru numit numărul gradelor de libertate al tabelului analizat. Prin numărul gradelor de libertate se înțelege numărul căsuțelor interioare ale unui tabel cu totalurile date, care se pot completa după voie.

Iată un extras din tabelul cu valorile critice ale lui

Nr.grade de libertate(ν)	P = o,o5	P = o,ol
1	3,841	6,635
2	5,991	9,21o
3	7,815	11,345
4	9,488	13,277
5	11,o7o	15,o68
6	12,592	16,812
7	14,o67	18,475
8	15,5o7	2o,o9o
9	16,919	21,666
lo	18,3o7	23,2o9

Pentru un grad de libertate, diferenţa devine semnificativă la nivelul o,o5 dacă χ^2 depăşeşte pe 3,841 şi la nivelul o,ol dacă depăşeşte 6,635.

Cîteva exemple vor lămuri foarte bine utilizarea testului.

a) Intr-un eşantion de loo persoane s-au găsit 45 de bărbaţi şi 55 femei. Se poate afirma că în populaţie sînt mai multe femei decît bărbaţi ?

Avem deci frecvenţele empirice:

$$n_1 = 45; \quad n_2 = 55$$

care se cer comparate cu frecvenţele teoretice:

$$m_1 = 5o ; \quad m_2 = 5o$$

obţinute în ipoteza coincidenţei numerice a indivizilor de sexe diferite. Există un singur grad de libertate, căci dacă fixăm o frecvenţă, cealaltă rezultă din diferenţa la total.

$$\chi^2 = \frac{(45-5o)^2}{5o} + \frac{(55-5o)^2}{5o} = \frac{25}{5o} + \frac{25}{5o} = \frac{5o}{5o} = 1$$

$$\chi^2 = 1 < \chi_0^2(P = o,o5) = 3,841$$

Deci ipoteza nulă nu poate fi respinsă, adică nu se poate spune că în populație ar fi mai multe femei decît bărbați.

Observație. Trebuie subliniat că variabila χ^2 face apel la frecvențele absolute, nu la cele relative sau procente. Dacă în cazul de mai sus s-ar multiplica toate cifrele cu 10 (adică dacă eșantionul ar fi fost de 1000, cu 450 bărbați și 550 femei), structura sa ar fi fost aceeași dar valoarea lui χ^2 ar fi fost 10, în loc de 1, depășind valoarea critică chiar și pentru nivelul 0,01.

b) Conform unor date existente, structura socială a unei colectivități este: 25% agricultori, 50% muncitori, 10% intelectuali și 15% alte categorii sociale. Intr-un eșantion de 1000 de persoane s-au obținut frecvențele:

Agricultori	Muncitori	Intelectuali	Alții	Total
230	470	135	165	1000

Este eșantionul bine construit ?

Conform ipotezei frecvențele teoretice sînt: 250, 500, 100 și 150.

$$\chi^2 = \frac{(230-250)^2}{250} + \frac{(470-500)^2}{500} + \frac{(135-100)^2}{100} + \frac{(165-150)^2}{150}$$

$$= \frac{400}{250} + \frac{900}{500} + \frac{1225}{100} + \frac{225}{150} = 1,6+1,8+12,25-1,5 = 17,15$$

Numărul gradelor de libertate fiind 3, rezultă:

$$\chi^2 = 17,15 > \chi_0^2 (P = 0,01) = 11,345$$

Deci valoarea este semnificativă, ceea ce înseamnă că diferența de structură dintre eșantion și populație nu poate fi pusă pe seama fluctuațiilor de eșantionare, ci pe o procedură distorsionantă.

c) Intr-o cercetare se urmărește detectarea unei asocieri între vîrstă și atitudinea față de o anumită problemă. Clasele caracteristicilor și datele ipotetice sînt prezentate în tabelul 1.

Vîrsta \ Atitudinea	Favorabilă	Neutră	Nefavorabilă	Total
Tineri	1oo	8o	2o	2oo
Adulţi	8o	1oo	2o	2oo
Vîrstnici	2o	2o	6o	1oo
TOTAL	2oo	2oo	1oo	5oo

Tabelul 1. Frecvenţe empirice

Pentru stabilirea unei eventuale asocieri, comparăm tabelul cu frecvenţe reale (empirice) cu cel de independenţă (frecvenţele teoretice). Se observă imediat o oarecare concentrare a indivizilor pe diagonala principală. Se ridică întrebarea dacă acest lucru poate sau nu să se datoreze numai fluctuaţiilor de eşantionare.

Vîrsta \ Atitudinea	Favorabilă	Neutră	Nefavorabilă	Total
Tineri	8o	8o	4o	2oo
Adulţi	8o	8o	4o	2oo
Vîrstnici	4o	4o	2o	1oo
TOTAL	2oo	2oo	1oo	5oo

Tabelul 2. Tabelul de independenţă, construit cu totalurile tabelului 1.

Se calculează:

$$\chi^2 = \frac{(1oo-8o)^2}{8o} + \frac{(8o-8o)^2}{8o} + \frac{(2o-4o)^2}{4o} + \frac{(8o-8o)^2}{8o} + \frac{(1oo-8o)^2}{8o} +$$

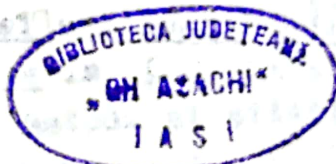
$$+ \frac{(2o-4o)^2}{4o} + \frac{(2o-4o)^2}{4o} + \frac{(2o-4o)^2}{4o} + \frac{(6o-2o)^2}{2o} =$$

$$= \frac{400}{80} + 0 + \frac{400}{40} + 0 + \frac{400}{80} + \frac{400}{40} + \frac{400}{40} + \frac{400}{40} + \frac{1600}{20} =$$

$$= 5 + 10 + 5 + 10 + 10 + 10 + 80 = 140$$

Numărul gradelor de libertate este 4. La acest număr, mărimea χ^2 obţinută este cu mult mai mare decît valorile critice din tabel, deci asocierea este semnificativă.

<u>Observaţie</u> . Ne reamintim că mărimea χ^2 a fost utilizată la calcularea coeficienţilor de asociere. Este clar că, dacă e vorba de o cercetare selectivă, primul pas ce trebuie făcut este cel de mai sus, adică testarea semnificaţiei asocierii. Abia apoi, cu aceeaşi mărime χ^2, se va trece la calcularea coeficientului de asociere, despre care se va putea spune, la rîndul său, că este semnificativ).



B I B L I O G R A F I E

ALUAS,I., DRAGAN, I.(1971) Sociologia franceză contemporană. Antologie, Ed.Politică, Bucureşti

BARBAT,Al. (1972) Teoria statisticii sociale, Ed.did. şi ped., Buc.

BIJI,M., BIJI, E. (1979) Statistică teoretică, Ed.did. şi ped., Buc.

BLALOCK, H.(1973) Introduction à la recherche sociale, Duculot, Gembloux

BOUDON,R., LAZARSFELD,P. (1965) Le vocabulaire des sciences sociales; Concepts et indices, Mouton, Paris

BOUDON,R., LAZARSFELD,P. (1966) L'Analyse empirique de la causalité, Mouton, Paris

BOUDON,R (1967) L'Analyse mathématique des faits sociaux, Blon, Paris

BOUDON,R. (1971) Les mathématiques en sociologie, P.U.F, Paris

BOUDON,R. (1973) L'Inégalité des changes, Armand Colin, Paris

BOURDIEU, P. et al. (1968) Le métier de sociologue, Mouton, Paris

BOURDIEU, P. et al.(1971) "Abdicaţiile empirismului" în I.Aluaş şi I.Drăgan (1971)

CAPLOW,Th. (197o) L'Enquête sociologique, Armand Colin, Paris

CHELCEA,S. (1975) Chestionarul în investigaţia sociologică,Ed. ştiinţifică şi enciclopedică, Bucureşti

CHELCEA,S. (1982) Experimentul în psihosociologie, Ed.ştiinţifică şi enciclopedică, Bucureşti

CHELCEA,S. (coordonator (1985) Semnificaţia documentelor sociale, Ed.ştiinţifică şi enciclopedică, Bucureşti

COSTEA,St. et al. (1983) Sociologia românească contemporană, Ed. ştiinţifică şi enciclopedică, Bucureşti

DURKHEIM,E.(1974) Regulile metodei sociologice, Ed.ştiinţifică, Bucureşti

DUVERGER,M. (1964) Méthodes des sciences sociales,P.U.F., Paris

FESTINGER,L.,KATZ,D. (1963) Les méthodes de recherche dans les sciences sociales, P.U.F., Paris

GEORGESCU,Fl. (1979) Indrumător pentru cercetarea sociologică în cultura fizică, Ed.Sport-Turism, Bucureşti

GRAWITZ,M. (1972) Méthodes des sciences sociales, Dalloz, Paris

HERSENI, T. (1982) Sociologie, Ed.ştiinţifică şi enciclopedică, Bucureşti

LAZARSFELD,P. (1966) L'Interprétation des relations statistiques comme procédure de recherche", în R.Boudon şi P. Lazarsfeld (1966)

LAZARSFELD,P. (1970) Philosophie des sciences sociales, Gallimard, Paris

LAZARSFELD,P. (1971) Qu'est-ce que la sociologie, Gallimard, Paris

LOETHER,H., J.McTAVISH,D.G. (1976) Descriptive and Inferential Statistics, Allyn and Bacon, Boston

MARGINEAN,I. (1982) Măsurarea în sociologie, Ed.ştiinţifică şi enciclopedică, Buc.

MERTON,R.K. (1972) Influenţa cercetării empirice asupra teoriei sociologice în Cunoaşterea faptului social, seria Teorie şi metodă şi ştiinţele sociale, vol.VIII, Ed. politică, Bucureşti

MIFTODE,V.(1982) Introducere în metodologia investigaţiei sociologice, Ed.Junimea, Iaşi

MIHOC, Gh. et al (1982) Modele de analiză statistică, Ed.ştiinţifică şi enciclopedică, Bucureşti

MIHU,A. (,1973) ABC-ul investigaţiei sociaologice, vol.II, Ed. Dacia, Cluj

MOSER,C.A- (1967) Metode de anchetă în investigarea fenomenelor sociale, Ed.ştiinţifică, Bucureşti

NICOLA,T. (1973) Experimentul social, Ed.Scrisul românesc, Craiova

NOELLE,E.(1966) Les sondages d'opinion, Minuit, Paris

PAGES,R.(1971) Experimentul în sociologie , în I.Aluaş, I.Drăgan (1971)

ROTARIU,T (1974) Influenţa factorilor mediu de provenienţă, origine socială şi pregătire şcolară asupra reuşitei la admiterea în învăţămîntul superior, în Studia, Series Sociologia.

SHILS,E. (1967) Tendinţele cercetării sociologice,în vol.Sociologia contemporană

SOROKIN,P.A. (1967) Diversitate şi unitate în sociològie , în vol. Sociologia contemporană

STAHL,H.H. (1974) Teoria şi practica investigaţiilor sociologice, vol.I. şi II, Ed.ştiinţifică şi enciclopedică, Buc.

STOETZEL,J., GIRARD,A. (1975) Sondajele de opinie publică. Ed. ştiinţifică şi enciclopedică, Bucureşti

SZCZEPANSKI,J.(1972) Noţiuni elementare de sociologie. Ed.ştiinţifică, Buc.

VLASCEANU,L. (1982) Metodologia cercetării sociologice. Ed.ştiinţifică şi enciclopedică, Bucureşti

YEOMANS,K.A. (1968) Statistics for the social scientist, vol.I şi II, Penguin Books, Harmondsworth

YULE,G.U., KENDAKK,M.G. (1969) Introducere în teoria statisticii, Ed.ştiinţifică, Bucureşti

ZAMFIR,C. et al. (198o) Dezvoltarea umană a întreprinderii, Ed. Academiei RSR, Bucureşti

x x x (1967) Sociologia contemporană, în seria Teorie şi metodă în ştiinţele socisle, vol.V., Ed.politică,Bucureşti

x x x (1973) Studentul şi societatea, Ed.Dacia, Cluj